What is the significance of test-score differences? Six psychometric meta-analyses on *g* loadings and IQ scores: The relation of inbreeding, visual impairment, schizophrenia, epilepsy, hearing impairment, and giftedness with general intelligence

Dennys B. Franssen

Master thesis
Supervised by dr. Jan te Nijenhuis
Faculty of Industrial and Organizational psychology
University of Amsterdam
May 2010

Acknowledgement

I would like to thank Dr. Jan te Nijenhuis, Dr. Welko Tomic, Dr. Henk van der Flier, Dr. Arne Evers, Dr. Santo Di Nuovo, Dr. Heiner Rindermann, Patricia Studeny MSc, Shaul Shalvi MSc, Rosina van Bloois MSc, Lise-Lotte Geutjes MSc, Birte Grimen MSc , Rionne van Rijk MSc, Stephanie Kämper MSc, Mr. Frits Kruijt, and Mr. Michael Mingroni for their enthusiastic support and helpful guidance on my project.

Table of contents

Abstract

Many studies show that *g*-loaded test scores can increase and decrease over time as a result of training, treatment, and educational interventions. We focus on the question to what extent increases or decreases in test scores also are increases or decreases in *g*. Based on a review of a large number of highly varied empirical findings we hypothesize that with regard to correlations between a vector of *g* loadings and a second vector a value of +1.00 means that variation in scores on the second variable is caused by biological factors, a value of –1.00 means that variation is caused by non-biological factors, and a value close to zero means that variation is caused by both biological and non-biological factors with roughly comparable effects.

Five psychometric meta-analyses (MAs) were performed to test these premises. A sixth MA was performed to validate a simplified method in which Wechsler composite scores are used instead of subtest scores. We predicted a strong positive correlation between vectors of lowered IQ scores following inbreeding, and vectors of *g* loadings; a strong negative correlation between lowered IQ scores following visual impairment and hearing impairment in children on average aged < 13 years, on the one hand, and the vectors of *g* loadings on the other hand; and strong positive correlations between IQ score differences following schizophrenia, epilepsy, and giftedness on the one hand, and the vectors of *g* loadings on the other hand.

Confirming our hypotheses, true correlations of .84 on inbreeding (K = 4; total N = 1,783); -.72 on visual impairment in children (K = 6; total N = 363); -.69 on hearing impairment in children (K = 2; total N = 1,287) were found. Contrary to our hypothesis, the exploratory MA on schizophrenia showed an uncorrected correlation of -.50 (K = 5; total N = 315). The exploratory MA on epilepsy showed an uncorrected correlation of .44, which is mildly supportive of our hypothesis (K = 7; total N = 445). The MA using a simplified method yielded an almost identical correlation as the previous more extensive one, implicating the feasibility of the new technique. Taken together, the findings of inbreeding, visual impairment, and hearing impairment increase the likelihood of the hypothesized link between *g* loadings and a dimension of biological causation. Limitations of the theory and implications for interventions aimed at increasing *g* are discussed.

Scores on cognitive tests provide the best general predictor of accomplishments in education, job training, and work. As a result, cognitive tests are widely used for selection and placement in organizations, and increasingly also in educational settings. However, many studies have shown that scores on cognitive tests are prone to change. IQ scores can increase by means of training (Caruso, Taylor, & Detterman, 1982; Ericsson & Lehmann, 1996; Nelson, Westhues, & MacLeod, 2003; Ramey, Bryant, & Suarez, 1985; Swanson & Lussier, 2001), or retesting (te Nijenhuis, van Vianen, & van der Flier, 2007), and decrease as a result of infectious disease (Alcock & Bundy, 2001), inbreeding (Schull & Neel, 1965; te Nijenhuis, Tomic, & Franssen, 2009), and severe malnutrition (Grantham-McGregor, Ani, & Fernald, 2001; Nwuga, 1977; Richardson, Birch, & Hertzig, 1973). In addition, many studies demonstrated intergenerational score gains on cognitive tests over the last half century. This secular increase in test scores, termed Flynn effect, has been reported in countries on all continents (Flynn, 2006). The question to which extent these score gains and losses on *g*-loaded tests represent permanent changes in cognitive ability has occupied researchers for decades. The differences in IQ test scores imply that intelligence is a variable that can be manipulated or trained. This paper contributes to the important discussion as to which interventions raise *g* and which do not.

The correlations between a number of variables and *g* loadings were examined by carrying out two full psychometric MAs on inbreeding and visual impairment, and three exploratory psychometric MAs on hearing impairment, epilepsy, and schizophrenia, respectively. The results should indicate which effects are associated with a true increase in general mental ability and which effects produce only "hollow" gains in test scores. A sixth meta-analysis was performed to test whether a simplified procedure requiring less data replicates the findings from a previous meta-analysis by te Nijenhuis, de Pater, van Bloois, and Geutjes (2009), using the same dataset on giftedness.

General Intelligence (*g*)

A well-established empirical finding—the manifold of positive correlations among measures of various mental abilities—is putative evidence of a general factor in all of the measured abilities. The method of factor analysis makes it possible to determine the degree to which each of the variables is correlated (or loaded) with the factor that is common to all the variables in the analysis. Spearman termed this *g* to represent a general factor that is manifested in individual differences on all mental tests, regardless of content (Jensen, 1998, p. 18). Spearman's *g* is best understood as a measure of cognitive complexity (Gottfredson, 1997), and is usually defined operationally as the loading on the first unrotated factor in a principal-axis factor analysis of a varied set of IQ tests (Jensen & Weng, 1994). Thus, tests demanding higher

cognitive complexity are high on *g* (have high *g* loadings), and tests demanding lower cognitive complexity are low on *g* (have low *g* loadings).

As noted, an interesting question is to what extent general mental ability can be increased or decreased and which effects will produce only "hollow" gains in test score. Jensen (1998) argued that training effects are most clearly manifested at the lowest level of Carroll's hierarchy, particularly on specific tests that most resemble the trained skills. One hierarchical level higher, the training effect is still evident for certain narrow abilities, depending on the nature of the training. However, the gain virtually disappears at the level of broad abilities and is altogether undetectable at the highest level, *g*. This implies that the transfer of training effects is strongly limited to tests or tasks all of which are dominated by one particular narrow skill or ability. Hence, there is virtually no transfer to tasks dominated by different narrow abilities, and none at the level of *g* itself. Thus any increase in narrow abilities or test-specific ability is independent of *g*.

Test-specific ability is defined as that part of a given test's true-score variance that is not common to any other test; i.e., it lacks the power to predict performance on any other tasks except those that are highly similar. Gains on test specificities are therefore not generalizable, but rather are "*empty*" or "*hollow*". Only the *g* component is highly generalizable. Jensen (1998, ch. 10) gives various examples of empty score gains, including a detailed analysis of the Milwaukee Project, which claimed very large increases in IQ score. However, Jensen's analysis indicates that there was no increase in *g*. Another example of empty score gains is given by Christian, Bachnan, and Morrison (2001) who state that increases due to schooling show very little transfer across domains.

Hierarchical Intelligence Model

Jensen (1998) hypothesized that scores on IQ batteries are best described by hierarchical intelligence models, such as Carroll's (1993) three-stratum hierarchical factor model of cognitive abilities. At the highest level of the hierarchy (stratum III) is general intelligence or *g*. One level lower (stratum II) is occupied by the broad abilities of Fluid Intelligence, Crystallized Intelligence, General Memory and Learning, Broad Visual Perception, Broad Auditory Perception, Broad Retrieval Ability, and Broad Cognitive Speediness or General Psychomotor Speed. One level lower still (stratum I) comprises the narrow abilities, including Sequential Reasoning, Quantitative Reasoning, Verbal Abilities, Memory Span, Visualization, and Perceptual Speed. The lowest level of the hierarchy consists of large numbers of specific tests and subtests. Some tests, despite seemingly very different formats, have empirically demonstrated to cluster into one narrow ability (Carroll, 1993).

Method of Correlated Vectors (MCV)

The question still remains as to which variables are associated with an increase or decrease in general intelligence. The MCV is a means of identifying variables that are associated with Spearman's *g*, the general factor of mental ability. This method involves calculating the correlation between: (a) the column vector of the *g* factor loadings of the subtests of an intelligence test or similar battery, and (b) the column vector of the relation of each of those same subtests with the variable in question. When the latter variable is dichotomous, the relations are usually calculated in terms of an effect size. When the latter variable is continuous (or nearly so), the relations are usually calculated in terms of a correlation coefficient (Ashton & Lee, 2005).

Although little has been written about the distribution of the values of the correlation between a *g* vector and a second vector, a clear picture has emerged from individual studies and meta-analyses (see Table 1). First of all, applying the MCV to biological variables such as head size, brain volume, brain's gray matter, brain's evoked potential, brain glucose metabolic rate, peripheral nerve conduction velocity, brain pH, body symmetry, giftedness, hybrid vigor, and inbreeding results in high positive correlations. After applying the statistical corrections typically carried out in psychometric meta-analysis (Hunter & Schmidt, 1990) it is not unlikely that the true correlation between *g* loadings and most of these biological variables approaches +1.00.

Second, a psychometric meta-analysis of correlations between vectors of *g* loadings and vectors of test-retest score gains based on a very large sample yielded a true correlation of -1.00 (te Nijenhuis, van Vianen, & van der Flier, 2007): the non-biological variable showing a perfect negative correlation with *g*. An exploratory study on learning potential in South-Africa (te Nijenhuis et al., 2007) reported a correlation of -.39 between score gains and the magnitude of *g* loadings of the items of Raven's Progressive Matrices. Correction for unreliability would probably yield a correlation of about -.80. Braden's (1989) study on the IQ scores of the non-genetic deaf found strong negative correlations between *g* loadings and the score difference between hearing and deaf groups. These negative correlations argue that the score gains or score differences are "hollow", that is, they are non-biological and do not represent true gains (or differences) in *g*.

Third, Jensen (1998, p. 320-321) was the first to ask the question whether the secular increase in test scores (the Flynn effect) is also correlated with *g* loadings. He reported data on four test batteries and found that these test's *g* loadings were not highly correlated with the amount of change. Subsequently, seventeen studies have examined whether secular trends are related to *g* (e.g., see Colom, Juan-Espinosa, & García, 2001; Flynn 1999ab, 2000; Must, Must, & Raudik, 2003; te Nijenhuis, & van der Flier, 2007; Wicherts et al., 2004) and have produced

conflicting results. A psychometric meta-analysis based on a very large sample which reported correlations between *g* loadings and standardized score gains on all studies having seven or more subtests yielded a true correlation of -.33; correction for statistical artifacts explained all the variance in the data points (te Nijenhuis & van der Flier, submitted, see Table 1). This correlation implies that less than half of the Flynn effect constitutes a real gain on *g*: the gain on non-*g* abilities being stronger than the gain on *g* suggesting that biological causes are less of an influence than non-biological causes. Therefore it is likely that the secular IQ gains only partially reflect a functional increase of real-life problem solving (general mental) ability.

Table 1

Various Studies on the Correlation Between a g Vector and a Second Vector

study	variable	*r*	*N*
Biological variables			
Jensen (1994)	head size	.64	286
Wickett, Vernon, & Lee (1994)	brain volume	.65	80
Schoenemann (1997)	brain volume	.51	72
	brain's cortical gray matter	.66	72
Schafer (1985)	brain's evoked potential habituation index	.77	52
Eysenck & Barrett (1985)	brain's averaged evoked potential	.95	219
Haier, Siegel, Tang, Abel, & Buchsbaum (1992)	brain's glucose metabolic rate	.79	8
Vernon (1992, 1993)	peripheral nerve conduction velocity	.44	85
Rae et al. (1996)	intercellular brain pH	.63	42
Colom, Jung, & Haier (2006)	brain gray matter	.82	23
	brain gray matter	.36	25
Lee et al. (2006)	brain activity	.61	36
Prokosch, Yeo, & Miller (2005)	body symmetry	.98	78
Schull & Neel (1965)	inbreeding	.79	865
Badarudozza & Afzal (1993)	inbreeding	.83	50
Nagoshi & Johnson (1986)	hybrid vigor	.52	2,096
Te Nijenhuis, de Pater, van Bloois, & Geutjes (2009)	gifted	**1.01**[1]	4,823
	mentally retarded	**.74**[1]	2,729
te Nijenhuis & Jongeneel-Grimen (2007)	heritability coefficients	**1.01**[1]	2,590
Mix biological/non-biological variables			
te Nijenhuis & van der Flier (submitted)	Flynn effect gains	**-.33**[1]	12,732
Non-biological variables			
te Nijenhuis et al. (2007)	test-retest gains	**-1.00**[1]	26,990
	learning potential training gains	-.39	95
te Nijenhuis & Jongeneel-Grimen (2007)	headstart gains	**-.80**[1]	602
te Nijenhuis & Jongeneel-Grimen (2007)	adoption gains	**-1.06**[1]	664

Braden (1989)	IQ scores of non-genetic deaf	-.76	325

Note. n.r. = not reported or could not be obtained. Many of the correlations were taken from Jensen (1998), but the authors of the original studies are listed in the Table. Schoenemann (1997) is cited in Jensen (1998, p. 147); sample sizes are not reported by Jensen and were taken from Schoenemann's dissertation.

Haier et al. (1992) show that there is an inverse relationship between brain glucose metabolic rate and psychometric measures of intelligence. A negative correlation is reported and we reversed the sign. Colom et al. (2006) report a collection of 28 correlations (Table 3) and 26 correlations (Table 5) on brain gray matter yielding the average correlation presented in the present Table. Lee et al. (2006) report in their Table 2 data on the activity in several brain regions. The average value of the sixteen correlations is reported in the present Table. Prokosch et al. (2004) report data on IQ scores and body symmetry. They also report the association between the rank-order of *g* loadings of five cognitive tests and its body symmetry association. We used their data to compute the rank-order correlation between rank-ordered *g* loadings and body symmetry association, which is $r_s = .98$. Schull & Neel (1965) tested 865 children from consanguineous marriages and 989 children from non-consanguineous marriages. Jensen (1983) uses the same data. Badaruddozza & Afzal (1993) tested 50 inbred and 50 non-inbred control children. Braden (1989) reports the correlation of the differences in IQ scores between normal and hearing-impaired individuals and *g* loadings. Braden reports a median $r = -.76$ for six studies, but the three largest studies are criticized by Isham & Kamin (1993). We take the $r = -.76$ as an estimate of the mean correlation for the remaining three studies (combined $N = 325$). [1] These correlations are based on meta-analyses and are corrected for artifacts.

These findings suggest the following theory: If the correlation between the *g* vector and a second vector is close to +1.00, variation in scores on the variable is caused by biological factors; If the correlation is close to -1.00, the variation in scores is caused by non-biological factors; If the correlation is close to 0.00, the variation is caused by both biological and non-biological factors. The last possibility also suggests that *g* and non-*g* skills play equal roles in effects.

Technically, a correlation of +1.00 or -1.00 implies that the variation in scores on specific variables can be perfectly predicted using *g* loadings, while a correlation of 0 implies that *g* loadings are useless for predicting the variation in the scores on the variables. However, we hypothesize that when these correlations are obtained by using the MCV additional information can be gleaned, namely the degree to which variation in scores on the specific variables is caused by biological and non-biological factors. Not only do very high positive or negative correlations yield important information, but correlations close to zero do as well.

Research Questions

Inbreeding depression: the biological mechanism explicated

It is undisputed that the brain is a product of biological evolution. In the course of human evolution there has been directional selection for increased brain size. During five million years of human evolution the human brain has tripled in size, a true explosion. As there are no discernible advantages, yet many disadvantages to increased brain size, natural selection must have acted directly upon the greater capacity for complex behaviors that promoted survival (Jensen, 1983). In the context of genetical theory of evolution, directional selection for a trait over many generations gradually increases the genetic dominance of those alleles which most enhance the advantageous phenotypic expression of the trait. The proportion of dominant alleles affecting a polygenic trait in a population is gradually increased by the fact that positive directional selection decreases those alleles whose combinations have strictly additive effects on the expression of the trait (Fisher, 1930). As inbreeding increases the proportion of homozygosity, the probability that paired alleles both will be recessive increases as well. To the extent that there is dominance to a trait inbreeding will lower the mean of that trait relative to the mean of a non-inbred but otherwise comparable population. This phenomenon is known as inbreeding depression. The theory of the genetic mechanisms responsible for the effects of inbreeding on polygenic traits is thoroughly explicated by Crow and Kimura (1970) and Jensen (1978, pp. 78-90). The fact that the effect of inbreeding depression is a hundred percent genetically leads to the following hypotheses:

(1) the true correlation between score differences between inbred groups and average groups and the magnitude of *g* loadings is strongly positive in sign;
(2) the true correlation between score differences between inbred groups and control groups and the magnitude of *g* loadings is strongly positive in sign.

Visual and Hearing impairment: Cumulative Deficit Hypothesis explicated

Group differences in mental scores are rather the rule than the exception (Berry, 1966; Jensen, 1980; Lynn, 1982, 1988, 1997; Lynn & Vanhanen, 2002; Ogbu, 1994; Reynolds & Murdoch-James, 1994; Shuey, 1966; Wright, Taylor, & Ruggiero, 1996; Zeidner, 1987). Cumulative deficit is a hypothesis concerning the cause of lower mental scores of groups considered environmentally deprived (Jensen, 1974). It presupposes a progressive decrement in test scores, relative to population norms, as a function of age. One of the most hotly debated issues pertaining to this paradigm is the group differences in IQ scores between Blacks and Whites in the US. Cumulative deficit argues that the deleterious effects of environmental and

cultural deprivation cumulate over time and lead to the 18 IQ points (1.2 *SD*) difference in mental scores repeatedly found between Blacks and Whites (Jensen, 1998; Rushton, 2000, 2001).

Visual impairment provides a means of testing the hypothetical assumptions underlying cumulative deficit through a natural experiment. Due to their lack of vision, the visually impaired grow up in a world deprived of visual stimuli. It is hypothesized that the blind and partially sighted experience a similar kind of environmental deprivation as do Black children; a poor and non-stimulating environment, negatively affecting mental scores. Of course, the environment of visually impaired is not exactly the same as the environment of Black children. However, it does not seem unreasonable to assume that the environment of the visually impaired is much more deprived than the environment of Black children. If the severe deprivation experienced by the visually impaired does not affect mental ability, it seems highly unlikely that the arguably less deprived environment of Black children is an important source of their lower mental ability scores.

Te Nijenhuis, van Rijk, and Kämper (2007) investigated the influence of visual deprivation on mental scores, by performing a meta-analysis on IQ scores of the visually impaired. They found that age has a substantial impact on the mean IQ of the visually impaired. The mean IQ score of visually impaired children aged 4-12 years was 11.9 IQ points lower than the IQ mental score of children and adults aged 13-79 years. This strongly contradicts expectations based on the cumulative deficit hypothesis, which would predict lower IQ scores for the older children and adults, because they experienced more years of deprivation. It appears that the low IQ of young blind children is caused by non-genetic factors, which, leads to the following hypothesis:

(3) the true correlation between score differences of groups of visually impaired children with an average age < 13 years and average groups and the magnitude of g loadings is strongly negative in sign.

Likewise, deaf children could also be considered environmentally deprived. The theoretical framework behind the hypothesis on visually impaired children is tested exploratorily on samples of hearing impaired children. It appears that the low IQ of young deaf children is caused by non-genetic factors, which, leads to the following hypothesis:

(4) the true correlation between score differences of groups of hearing impaired children with an average age < 13 years and average groups and the magnitude of g loadings is strongly negative in sign.

Exploratory analyses: Schizophrenia and Epilepsy

Research studying the effects of epilepsy on cognitive abilities produced inconsistent results. Some researchers report that children and adolescents tend to score consistently lower than same age peers on measures of intelligence (Reichenberg & Harvey, 2007; Smith, Elliott, & Lach, 2002), whereas other researchers found that children with epilepsy had no measurable cognitive difficulties (Hauser & Hesdorffer, 1990).

In children and adolescents with schizophrenia lowered intelligence is often found, and cognitive functioning either may further deteriorate, remain stable, or even improve slightly. The processes leading to neuropsychological deficits in schizophrenia are poorly understood. While there is a consensus that neuropsychological deficits are central characteristics of schizophrenia, it is difficult to distinguish deficits that reflect abnormal development from those that reflect deterioration of acquired abilities. We speculate that the *g* factor will be affected by the brain damage that coincides with schizophrenia. Therefore, we expect:

(5) the true correlation between score differences of a schizophrenic group and a comparison group and the magnitude of the *g* loadings to be strongly positive in sign.

A decrease in intellectual functioning and impairment of cognitive abilities in patients with epilepsy has been described for many years. Several epilepsy-related factors, such as type, etiology, age at onset, localization, severity and duration of seizures, and heredity are generally believed to affect the IQ of patients with epilepsy. Lowered IQ scores have been reported in persons with generalized seizures as compared to those with partial or psychomotor seizures (Klove & Mathews, 1966) and in children with symptomatic epilepsy as compared to those with idiopathic epilepsy (Bourgeouis, Prensky, Palkes, Talent, & Busch, 1983). However, some researchers believe that most children with epilepsy are of normal intelligence, and do not deteriorate over time (Bourgeouis et al., 1983). The precise nature of the cognitive deficit associated with generalized epilepsy and epilepsy-related factors is not clear. The present study investigates lowered IQ scores in patients with epilepsy. We hypothesize that a biological factor underlies lowered IQ scores in patients with epilepsy, which leads to the following hypothesis:

(6) the true correlation between score differences of an epileptic group and a comparison group and the magnitude of the *g* loadings is strongly positive in sign.

Methodological experiment using giftedness data

The sixth MA was performed as a methodological experiment. Part of the dataset on giftedness of the MA by te Nijenhuis, de Pater, van Bloois, and Geutjes (2009) was used to

validate a new method. Instead of using seven subtests, the current MA was performed using Wechsler composite *Verbal* (*V*), *Performance* (*P*), and *Full Scale* (*FS*) IQ scores to compute *d* values and *g* loadings. Throughout this project we noticed that a relatively small percentage of studies report IQ scores on subtest level, while a much larger percentage of studies report Wechsler composite *V*, *P*, and *FS* IQ scores. te Nijenhuis et al. (2009) found a true correlation between score differences between a gifted group and an average group and the magnitude of *g* loadings of 1.01 (N = 4,823), with 57 % of the variance in the datasets explained by five statistical artifacts. We argue that when the current study using the simplified method replicates the outcomes of the previous one, this would validate the use of the new simplified method. This leads to the following hypothesis:

(7) when using the simplified MCV the true correlation between score differences of gifted and average groups and the magnitude of *g* loadings is strongly positive in sign.

General Method

Psychometric meta-analysis (Hunter & Schmidt, 1990) estimates what the results of studies would have been if all studies had been conducted without methodological limitations or flaws. The results of perfectly conducted studies would allow a clearer view of underlying construct-level relationships (Schmidt & Hunter, 1999). The goal of the present study is twofold. First, to provide reliable estimates of the true correlation between a number of variables and the magnitude of *g* loadings. These variables are: inbreeding, visual impairment, hearing impairment, schizophrenia, and epilepsy. Second, the meta-analysis on giftedness is a means of testing a new technique of combining psychometric meta-analysis with a simplified version of the method of correlated vectors. By reanalyzing the data of the meta-analysis performed by te Nijenhuis, de Pater, van Bloois, and Geutjes (2009) using composite Verbal, Performance, and Full Scale IQ scores instead of scores on subtest level, a new technique of investigating the true correlation between variables and the magnitude of *g* loadings is introduced.

In general, *g* loadings were computed by submitting a correlation matrix to a principal-axis factor analysis and using the loadings of the subtests on the first unrotated factor. In some cases *g* loadings were taken from studies where other procedures were followed; these procedures have been shown empirically to lead to highly comparable results. Finally, Pearson correlations between each of the four variables (score differences between an inbred group and an average group, a visually impaired group and an average group, a schizophrenic group and an average

group, and an epileptic group and an average group, and the *g* loadings) were computed.

There has been a discussion whether one should use Pearson *r* or Spearman's rho when applying the method of correlated vectors. The answer depends on whether one assumes an interval or an ordinal measurement level for IQ scores. Ranking of IQ scores can be seen as a way of categorizing intelligence levels on an ordinal scale. For instance, an IQ score of 150 indicates a higher level of intelligence compared to an IQ score of 75. However, the inference that an IQ score of 150 indicates a doubling in level of intelligence compared to an IQ score of 75 cannot be drawn.

In order to obtain our results, mean IQ scores were used to calculate the score differences between groups (*d*). Score differences have the characteristics of an interval scale: arithmetical operations can be conducted, and the effects (*d*) have values ranging from negative to positive. Thus, the choice for Pearson *r* or Spearman's rho depends on whether the underlying construct on which calculations are carried out are more important or the calculations themselves. Colom, Juan-Espinosa, Abad, and Garcia (2000) consider both Pearson *r* and Spearman's rho as suitable measures of the degree of relationship between two vectors. We decided to use Pearson *r* following earlier conducted meta-analyses using Pearson *r* in the method of correlated vectors (te Nijenhuis, van Vianen, & van der Flier, 2007; te Nijenhuis & Jongeneel-Grimen, 2007; te Nijenhuis, de Pater, van Bloois, & Geutjes, 2009). This has the advantage that the results of the present studies can be compared directly against those of the earlier studies.

General Inclusion Rules

For studies to be included in a MA four criteria had to be met: First, in order to obtain a reliable estimate of the true correlation between each of the four variables (inbreeding, visual impairment, schizophrenia, and epilepsy) and the *g* loadings, the cognitive batteries had to have a minimum of seven subtests. For the MA on giftedness the requirement was different: all studies reporting Wechsler composite *Verbal*, *Performance*, and *Full Scale* IQ scores were included in the MA on giftedness. Second, the IQ test had to be well-validated. Third, since studies with a test-retest effect would influence the 'true' correlation between *d* and *g* (see discussion below) – they were excluded. That is, studies using a counterbalanced design and the scores of the re-administration of an IQ battery within a test-retest design were not included. In a counterbalanced design, participants are administered two IQ batteries, X and Y, in different orders. Half of the participants take test X first, then test Y and vice versa. Finally, only studies published in English, Dutch, or German were used.

Test-retest Effects on 'true' g

If one takes the exact same test battery a second time the test-retest effect is, by definition, at 100% of its strength. The effect of taking a test twice is modest – only a few IQ points (Jensen, 1980; Kulik, Bangert-Drowns, & Kulik, 1984). Additional training can increase the size of the score gains. Kulik et al.'s MA on test preparation studies resulted in effect sizes on intelligence tests for practice and additional coaching of 0.25 *SD* and 0.51 *SD*, respectively. In addition, the true correlation between test-retest score gains and *g* loadings has been shown to be -1.00, based on a psychometric MA with a very large sample size (te Nijenhuis, van Vianen, & van der Flier, 2007). Te Nijenhuis et al. argue that the gains are linked to test-specific variance only, and not at all to the variance associated with general, broad, or narrow abilities.

There are occasions in which people take two comparable test batteries, such as the WISC and the WISC-R, or the WISC and the WAIS. The time elapsed between testing moments can vary from taking two tests directly, to decades separating testing moments. The latter is still regarded as test-retest bias. In some studies people also take test batteries that are non-comparable, i.e. constructed according to different psychometric principles, such as, the WISC-R and the Kaufman-ABC. The size of the test-retest effect is most strongly influenced by the degree of similarity between test batteries taken. When two non-comparable tests are taken, such as first the WISC-R and then the Kaufman-ABC, these tests both measure the *g* factor, but with different flavors (see Carroll, 1993). So, a number of the principles in the subtests of the WISC-R cannot be applied to the subtests of the K-ABC. Therefore, the test-retest effect will be weaker.

Overall, test-retest effects mask the theoretically expected true correlation of +1.00 for the biological variables, and -1.00 for the non-biological variables. Therefore, datasets with test-retest effects are excluded. However, in some cases scores on the first test were used for meta-analysis.

Corrections for Artifacts

Psychometric meta-analytical techniques (Hunter & Schmidt, 1990, 2004) were applied using the software package developed by Schmidt and Le (2004). Psychometric meta-analysis is based on the principle that there are artifacts in every dataset and that most of these artifacts can be corrected. In the present meta-analyses we corrected for five artifacts identified by Hunter and Schmidt (1990) that alter the value of outcome measures. These are: (1) sampling error, (2) reliability of the vector of *g* loadings, (3) reliability of the vector of a specific variable of theoretical interest (4) restriction of range of *g* loadings, and (5) deviation from perfect construct validity. In the present exploratory studies, using bare-bones meta-analytical techniques, we corrected for only one artifact, namely sampling error.

Correction for Sampling Error

In many cases sampling error explains the majority of the variation between studies, so the first step in a psychometric meta-analysis is to correct the collection of effect sizes for differences in sample size between the studies.

Correction for Reliability of the Vector of g Loadings

The values of *r (g × inbreeding), r (g × visual impairment), r (g × hearing impairment), r (g × schizophrenia), r (g × epilepsy), and r (g × giftedness)* are attenuated by the reliability of the vector of g loadings for a given battery. When two samples have a comparable *N*, the average correlation between vectors is an estimate of the reliability of each vector. Several samples that differed little on background variables were compared. For the comparisons using children we chose samples that were highly comparable with regard to age. Samples of children in the age of 3 to 5 years were compared against other samples of children who did not differ more than 0.5 year of age. Samples of children in the age of 6 to 17 years were compared against other samples of children who did not differ more than 1.5 year of age. For the comparisons of adults we compared samples in the age of 18 to 95 years.

Correlation matrices were collected from test manuals, books, articles, and technical reports. The large majority came from the U.S., but also from European countries, and a substantial number from Korea, China, Hong Kong, and Australia. This resulted in about 700 data points, which yielded 385 comparisons of *g* loadings of comparable groups from which to estimate the reliability for that group. To give an illustration of the procedure, van Haasen et al. (1986) report correlation matrices of the Dutch and the Flemish WISC-R for 22 samples in the age of 6-16 years. Samples of children in the age of 6 to 17 years were compared to other samples of children who do not differ by more than 1.5 years. Because the samples of children reported in van Haasen et al. (1986) were between 6 and 17 years only children were compared who did not differ more than 1.5 years. The *N*s in these samples were comparable. The resulting average correlation was .78 (combined N = 3,018; average N = 137).

A scatter plot of reliabilities against *N*s should show that the larger *N* becomes, the higher the value of the reliability coefficients, with an asymptotic function between *r* (*g* x *g*) and *N* expected. The curve that gave the best fit to the expected asymptotic function was selected. The logarithmic regression line resembled quite well the expected asymptotic distribution for reliabilities. However, because the extreme range on the X-axis resulted in a picture that is not informative, the regression line for *r* (*g* x *g*) and *N* is not reported. For the same reason Figure 1 is divided into three parts, each showing the scatter plot of reliability of the vector of *g* loadings and sample size for a specific range of *N*.

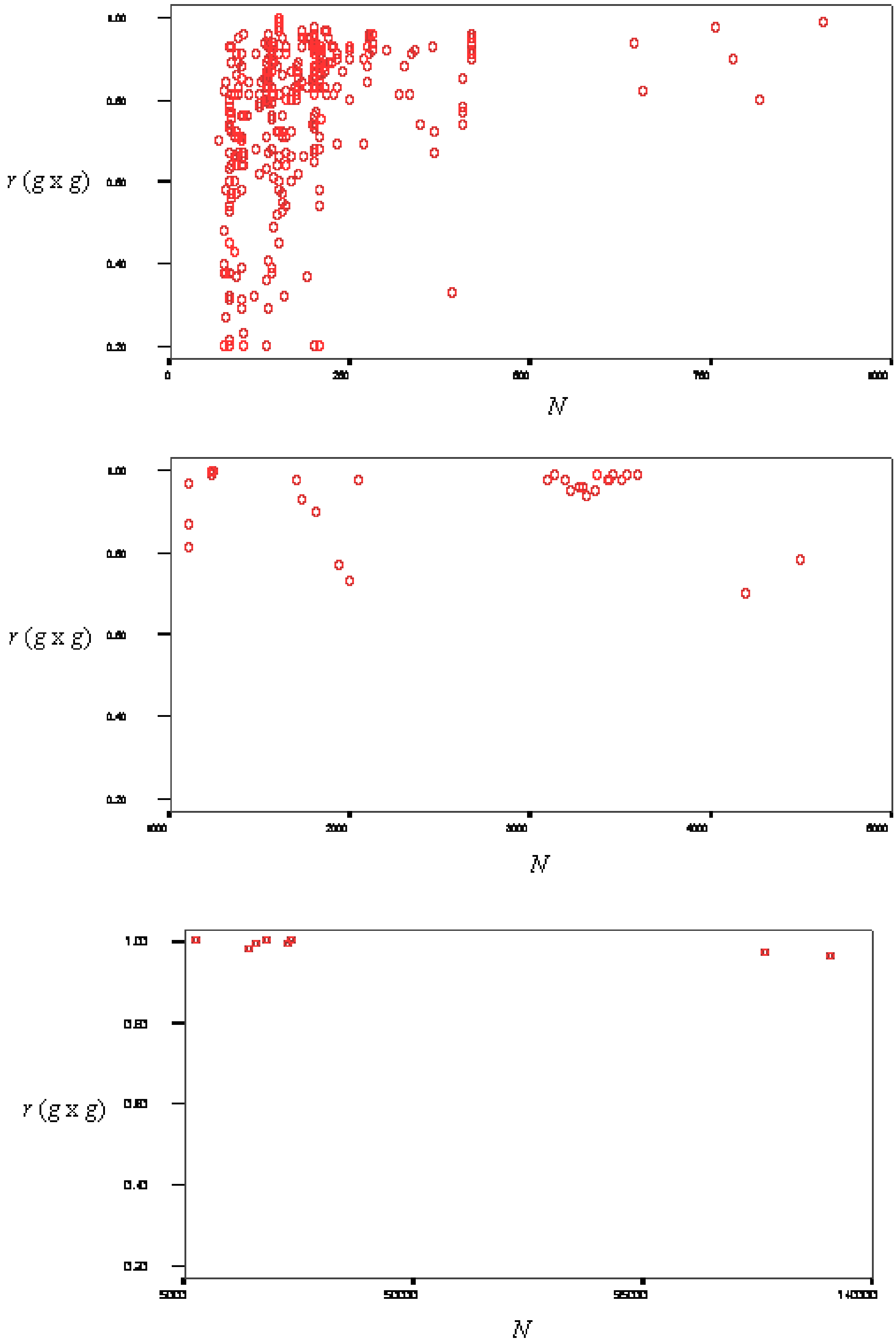

Figure 1

Three Scatter Plots of Reliability of the Vector of g Loadings and Sample Size Each for a Different Range of N

Correction for Reliability of the Vector of the Second Variable.

The values of *r (g × inbreeding), r (g × visual impairment), r (g × hearing impairment), r (g × schizophrenia), r (g × epilepsy),* and *r (g × giftedness)* are attenuated by the reliability of the vector of the second variable for a given battery. When two samples have a comparable N, the average correlation between vectors is an estimate of the reliability of each vector. The reliability of the vector of inbreeding, visual impairment, schizophrenia, epilepsy, and giftedness were each estimated using the present datasets, comparing samples that took the same test, and that differed little on background variables. For the comparisons using children we chose samples that were highly comparable with regard to age, and for the comparisons of adults we chose samples that were roughly comparable with regard to age.

Correction for Restriction of Range of g Loadings.

The values of *r (g × inbreeding), r (g × visual impairment), r (g × hearing impairment), r (g × schizophrenia), r (g × epilepsy), and r (g × giftedness)* are attenuated by the restriction of range of g loadings in many of the standard test batteries. The most highly g-loaded batteries tend to have the smallest range of variation in the subtests' g loadings. Jensen (1998, pp. 381-382) showed that restriction in the magnitude of g loadings strongly attenuates the correlation between g loadings and standardized group differences. Hunter and Schmidt (1990, pp. 47-49) state that the solution to variation in range is to define a reference population and express all correlations in terms of it. The Hunter and Schmidt meta-analytical program computes what the correlation in a given population would be if the standard deviation were the same as in the reference population. The standard deviations can be compared by dividing the standard deviation of the study population by the standard deviation of the reference group, that is $u = SD_{study}/SD_{ref}$. As references we used tests that are broadly regarded as exemplary for the measurement of intelligence, namely the various versions of the Wechsler tests for children and adults. The average standard deviation of g loadings of the various versions of the Wechsler Bellevue (W-B), Wechsler Preschool and Primary Scale of Intelligence (WPPSI), Wechsler Intelligence Scale for Children (WISC), Wechsler Intelligence Scale for Children–Revised (WISC-R), Wechsler Intelligence Scale for Children–Third Edition (WISC-III), and the Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV) from datasets from countries all over the world was 0.132. We used this value as our reference in the studies with children. The average standard deviation of g loadings of the various versions of the Wechsler Adult Intelligence Scale (WAIS), Wechsler Adult Intelligence Scale–Revised (WAIS-R), and the Wechsler Adult Intelligence Scale–Third Edition (WAIS-III) from datasets from countries all over the world was 0.107. This was used as the reference value in the studies with adults. In so doing, the SD of g loadings of all

test batteries was compared to the average SD in g loadings in the Wechsler tests for children and adults, respectively.

The Hunter and Schmidt meta-analytical program computes only the aforementioned four corrections. The observed correlation corrected for sampling error, unreliability of the vector of *g* loadings and the second vector, and range restriction is referred to as rho-4.

Correction for Deviation from Perfect Construct Validity.

The deviation from perfect construct validity in g attenuates the values of *r (g × inbreeding), r (g × visual impairment), r (g × hearing impairment), r (g × schizophrenia), r (g × epilepsy),* and *r (g × giftedness).* In making up any collection of cognitive tests, we do not have a perfectly representative sample of the entire universe of all possible cognitive tests. Therefore any one limited sample of tests will not yield exactly the same g as another such sample. The sample values of g are affected by psychometric sampling error, but the fact that g is very substantially correlated across different test batteries implies that the differing obtained values of g can all be interpreted as estimates of a "true" g. The values of *r (g × inbreeding), r (g × visual impairment), r (g × hearing impairment), r (g × schizophrenia), r (g × epilepsy),* and *r (g × giftedness)* are attenuated by psychometric sampling error in each of the batteries from which a *g* factor has been extracted.

The more tests and the higher their *g* loadings, the higher the *g* saturation of the composite score is. The Wechsler tests have a large number of subtests with quite high *g* loadings, yielding a highly *g*-saturated composite score. Jensen (1998, p. 90–91) states that the *g* score of the Wechsler tests correlates more than .95 with the tests' IQ score. However, shorter batteries with a substantial number of tests with lower *g* loadings will lead to a composite with somewhat lower *g* saturation. Jensen (1998. ch. 10) states that the average *g* loading of an IQ score as measured by various standard IQ tests lies in the +.80s. When this value is taken as an indication of the degree to which an IQ score is a reflection of "true" g, it can be estimated that a tests' *g* score correlates about .85 with "true" *g*. As *g* loadings represent the correlations of tests with the *g* score, it is most likely that most empirical *g* loadings will underestimate "true" *g* loadings; therefore, empirical *g* loadings correlate about .85 with "true" *g* loadings. As the Schmidt and Le (2004) computer program only includes corrections for the first four artifacts, the correction for deviation from perfect construct validity was carried out on the values of *r* (*g* × inbreeding), *r* (*g* × visual impairment), *r* (*g* × hearing impairment), *r* (*g* × schizophrenia), *r* (*g* × epilepsy), and *r* (*g* × giftedness), after correction for the first four artifacts. To limit the risk of overcorrection, we conservatively chose the value of .90 for the correction. The observed correlation corrected for sampling error, unreliability, range restriction, and imperfect construct validity is referred to as

rho-5.

Study 1: Inbreeding depression

To test whether there is a strong positive correlation between the magnitude of *g* loadings and IQ scores of inbred children, a psychometric meta-analysis was performed on all studies that reported IQ scores of at least seven subtests from children of consanguineous parentage. The majority of subjects in this MA were offspring from first-cousin marriages, which results in an inbreeding coefficient of .063. The coefficient of inbreeding (*f*) is the average probability over all gene loci that the same allele on both homologous chromosomes comes from the same ancestor (Crow & Kimura, 1970, pp. 64-65). Thus, if there is dominance to the alleles which enhance the phenotypic expression of the trait, inbreeding will lower the mean of the trait relative to the mean of a non-inbred but otherwise comparable population − the phenomenon known as inbreeding depression. The theory of the genetic mechanisms responsible for the effects of inbreeding on polygenic traits is thoroughly explicated by Crow and Kimura (1970) and Jensen (1978). The mean coefficient of inbreeding (*f*) in this MA was .047.

Method

Searching and screening studies. Starting point for the current meta-analysis is the dataset from te Nijenhuis, Tomic, and Franssen (2009) who investigated the relationship between the degree of inbreeding (*f*) and depression of IQ scores. Four methods were used to identify studies that contained IQ scores of inbred offspring. First, an electronic search for published research using PsycINFO, ERIC, MEDLINE, PiCarta, Academic search premier, Web of science, Google Scholar, and PubMed was conducted. Keywords used were inbred*, inbreeding*, incest*, consanguin*, cognitive, mental ability, intelligence, IQ, WISC, Wechsler, and combinations of these concepts (* is a truncation symbol to represent multiple spellings or endings; AND is a Boolean operator that combines search terms so that the search result contains all of the terms). Second, the reference lists of all significant articles were analyzed in search of additional studies. Third, cited reference searches were conducted using Web of Science, to search for articles citing significant articles. Last, several authors were asked for additional studies on the subject.

This procedure resulted in 46 articles, book chapters, and reports on the concurrent topics of inbreeding depression and mental ability. Only four studies met all criteria for inclusion in the MA, comprising all published research on the subject published in English-language research journals and books.

Specific criteria for inclusion. For a study to be included in the meta-analysis, three additional criteria had to be met: First, only empirical studies reporting an inbreeding coefficient

(*f*) were included. Second, the mean subtest scores had to be lower than the mean scores of the standardization sample of the IQ test. Finally, studies that reported additional variables known to influence mental performance were excluded. Application of these inclusion rules yielded four studies resulting in four correlations between *g* and score differences between an inbred group and an average group. The Israeli study by Cohen, Bloch, Flum, Kadar, and Goldschmidt (1963) was left out of the MA because (1) it was unclear what the seventh subtest, called Substitution, represents, so only six subtests were left, (2) we were unable to obtain any information on the Israeli WISC, the test administered in this study, and (3) no information was reported on the age of the inbred children.

Computation of score differences between an inbred group and a comparison group. Score differences between an inbred group and an average group (*d*) were computed by subtracting the mean score of the inbred group of the particular test in question from the mean score of the standardization group, and then dividing the result by the *SD* of the standardization group. The standardization group scores were obtained by computing a weighted average score matching the age range of the inbred group as closely as possible. The *g* loadings were obtained in the same way. The weighted average *g* loadings were computed matching the age range of the inbred group to the age range of the *g* loadings as close as possible. Psychometric meta-analytical techniques (Hunter & Schmidt, 1990, 2004) were applied to the resulting four *r* (*g* × inbreeding)s using the software package developed by Schmidt and Le (2004). In the present study we corrected for the five artifacts (mentioned above) that alter the value of outcome measures listed by Hunter and Schmidt (1990).

Three studies also reported a control group. A bare-bones MA was performed on these three studies. Score differences between an inbred group and the control group (*d*) were computed by subtracting the mean score of the inbred group of the particular test in question from the mean score of the control group, and then dividing the result by the *SD* of the *standardization* group.

The results of the studies on the correlation between *g* loadings and the score differences between inbred groups on the one hand, and standardization groups (*d*) and control groups (*d*) on the other hand are shown in Table 2. The results of the psychometric MA using the standardization groups are shown in Table 3. The results of the bare-bones MA using control groups are shown in Table 4.

Correction for reliability of the vector of inbreeding. The value of *r* (*g* × inbreeding) is attenuated by the reliability of the vector of inbreeding for a given battery. The reliability of the vector of inbreeding was estimated using the present datasets, comparing samples that took the

same test and that were comparable in regard to age and sample size. As an illustration of the procedure, the following rules were set in order to analyze studies that were highly comparable. First of all, only studies using the same test and the same version of this test were taken together. Second, studies containing less than a hundred participants were considered to be highly comparable as long as the difference in *N* between two studies was lesser than or equal to sixty. Third, studies containing more than a hundred participants were considered to be highly comparable as long as the difference in *N* between two studies was lesser than or equal than hundred-fifty. Fourth, the difference in average age of participants in separate studies was three years or less. Finally, the date of publication between two studies did not differ more than ten years.

A scatter plot of reliabilities against *N*s should reveal that the larger *N* becomes, the higher the value of the reliability coefficients, with an asymptotic function between *r* (*d* x *d*) and *N* expected. We checked to see which curve gave the best fit to the expected asymptotic function. Figure 2 shows the scatter plot of reliability of the vector of inbreeding depression and sample size, and the logarithmic curve that fitted optimally.

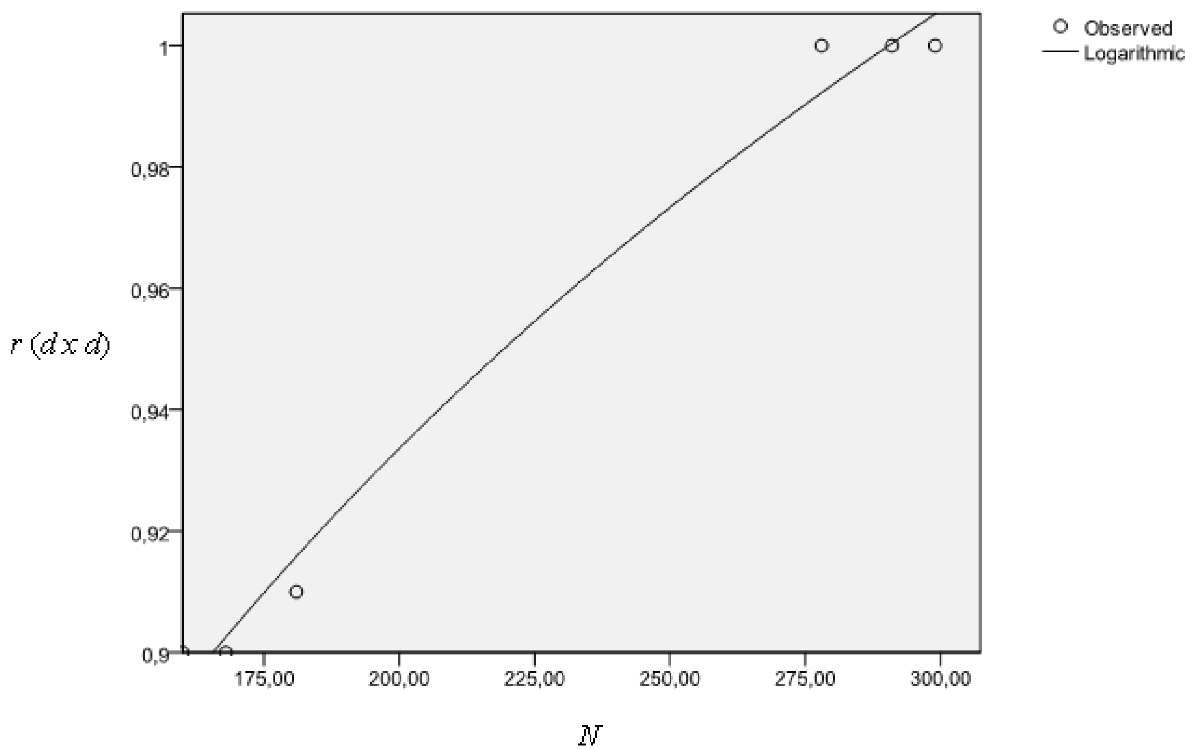

Figure 2
Scatter Plot of Reliability of the Vector of Inbreeding Depression and Sample Size and Regression Line

Results

The results of the studies on the correlation between *g* loadings and the score differences between inbred groups on the one hand, and average groups (*d*) and control groups (*d*) on the other hand are shown in Table 2. The Table gives data derived from four studies, with participants numbering a total of 2,349. It also lists the reference for the study, the cognitive ability test used, the correlation between *g* loadings and *d*, the sample size, and the mean age (and range of age). It is clear that the majority of the correlations are strongly positive.

Table 3 presents the results of the psychometric meta-analysis of the four data points. It shows (from left to right): the number of correlation coefficients (*K*), total sample size (*N*), the mean observed correlations (*r*) and their standard deviation (SD_r), the correlations one can expect once artifactual error from unreliability in the *g* vector, the inbreeding depression vector, and range restriction in the *g* vector have been removed (rho-4), and their standard deviation ($SD_{\text{rho-4}}$), and the true correlation one can expect when corrections for all five artifacts have been carried out (rho-5). The next two columns present the percentage of variance explained by artifactual errors (%VE), and the 80% confidence interval (80% CI). This interval denotes the values one can expect for rho in sixteen out of twenty cases.

The analysis of all four data points yields an estimated correlation (rho-4) of .27, with only 2% of the variance in the observed correlations explained by artifactual errors. However, Hunter and Schmidt (1990) state that extreme outliers should be left out of the analysis, because they are most likely the result of errors in the data. They also argue that strong outliers artificially inflate the *SD* of effect sizes and thereby reduce the amount of variance that artifacts can explain. Figure 3 shows the scatter plot of correlations *r* (*d* x *g*) against sample size. We choose to leave out one extreme outlier, namely Afzal (1988) with a value of *r* more than 17 *SD* beneath the average *r* of the final sample of three data points, using the *SD* of the final sample. This resulted in a value of the correlation (rho-4) of .76, a large decrease in the *SD* of rho-4 with 98%, and a fortyfold increase of the amount of variance explained in rho-4 by artifacts: 95% of the variance is now explained. Finally, a correction for deviation from perfect construct validity in *g* took place, using the conservative value of .90. This resulted in a value of .84 for the final estimated true correlation between *g* loadings and inbreeding depression.

Table 4 presents the results of the bare-bones MA of the three data points. It shows (from left to right): the number of correlation coefficients (*K*), total sample size (*N*), the true correlation (rho) and their standard deviation (SD_r) The last column present the percentage of variance explained by artifactual errors (%VE). The analysis of all three data points yields an estimated correlation (rho) of .03, with only 4% of the variance in the observed correlations explained by

artifactual errors.

Table 2

Studies of Correlations Between g Loadings and Inbreeding Depression

Reference	test	*r* (*d* x *g*)	*N*	*m* age (range)
inbreeding-standardization				
Afzal (1988)	WISC-R	-.73	566	10.5 (9.0-12.0)
Badaruddoza (2003)[a]	WISC-R	.63	868	8.3 (6.0-11.0)
Badaruddoza & Afzal (1993)[b]	WISC-R	.51	50	8.5 (6.0-11.0)
Schull & Neel (1965)[c]	WISC[d]	.50	865	8.6 (7.0-10.0)
inbreeding-control				
Afzal (1988)	WISC-R	-.24	566	10.5 (9-12)
Badaruddoza & Afzal (1993)	WISC-R	-.34	50	8.5 (6-11)
Schull & Neel (1965)	WISC[d]	.22	865	8.6 (7-10)

Note. [a]A weighted average was computed of first-cousins, first-cousins once removed, and second cousins. [b]Jensen (1983) reported a value of r = .83. [c]Jensen (1983) reported a value of r = .79. [d]The Japanese version of the WISC was used.

Table 3

Meta-analytical Results for the Correlation Between Inbreeding Depression and g Loadings after Corrections for Reliability, Restriction of Range, and imperfect Construct Validity Using Inbreeding-Standardization Comparisons

Predictor	*K*	*N*	*r*	SD_r	rho-4	$SD_{\text{rho-4}}$	rho-5	% VE	80% CI
Inbreeding depression[1]	4	2349	.25	.55	.27	.59	.30	2%	-.48-1.03
Inbreeding depression minus one outlier[2]	3	1783	.56	.06	.76	.01	.84	95%	.74-.78

Note. [1]Meta-analytical results for correlations between *g* loadings and *d* (inbreeding depression).[2]The study by Afzal (1988) is considered an extreme outlier and therefore removed. *K* = Number of correlations; *N* = Total sample size; *r* = mean observed correlation (sample size weighted); *SDr* = Standard deviation of observed correlation; rho = true correlation (observed correlation corrected for unreliability and range restriction); *SD*rho = Standard deviation of true correlation; %VE = Percentage of variance accounted for by artifactual errors; 80% CI = 80% credibility interval.

Table 4

Exploratory Bare-bones Meta-analytical Results for Correlations Between g Loadings and Inbreeding-Control Score Differences Using Inbreeding-Control Comparisons

Predictor	*K*	*N*	rho	SD_{rho}	%VE
Inbreeding-Control[1]	3	1,481	.03	.23	4%

Note. [1]Bare-bones meta-analytical results: Score differences between an inbred group, a matched control group, and *g* loadings. *K* = number of correlations; *N* = total sample size; rho = true correlation (observed correlation corrected for sample size); SD_{rho} = standard deviation of true correlation; %VE = percentage of variance accounted for by artifactual errors.

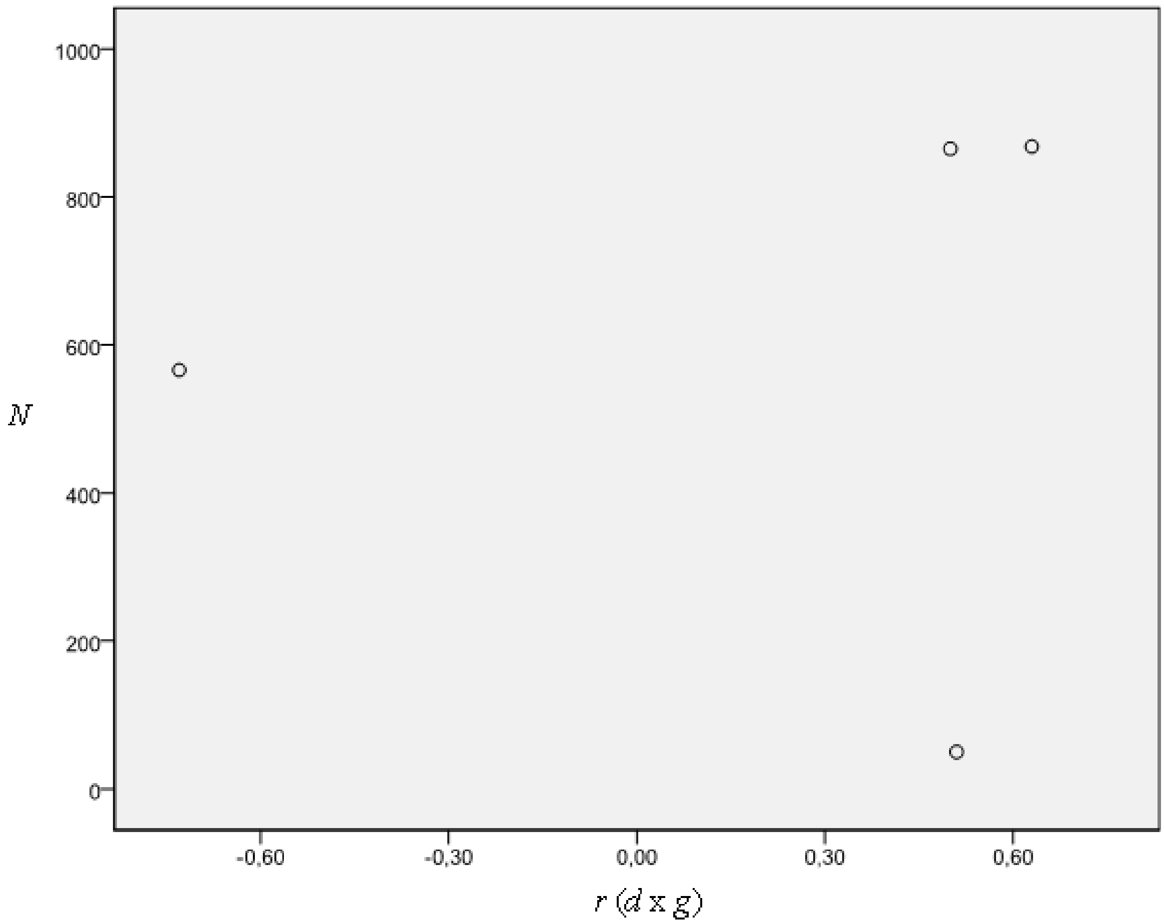

Figure 3

Scatterplot of Correlations (d x g) and Sample Size of the Variable Inbreeding depression Using Inbreeding-Standardization Comparisons

Conclusion

It is concluded that the data support the first hypothesis quite strongly: a correlation r_{gd} of +1.00 was expected and a value of +.84 is found. This is quite close to the predicted value in a meta-analysis based on a small number of studies. In sum, the findings are strongly in line with the findings on biological variables of te Nijenhuis and Grimen (2007), who showed the r_{gd} is +1.01 for heritability coefficients, based on a meta-analysis with a sample size of N = 2,590.

The second hypothesis is not supported. The bare-bones MA results in a rho of .03, while a value of +1.00 was expected.

Study 2: Visual Impairment

To test whether there is a –1.00 correlation between the magnitude of the *g* loadings of IQ subtests and visually impaired children under the age of 13.25 years a psychometric meta-analysis of all studies of visually impaired children that reported scores of at least seven subtests was performed. Visual impairment entails a severe limitation of visual capability and includes both partial sightedness and blindness (Bailey & Hall, 1990). In this meta-analysis the following definitions are used: Visual impairment is defined as having less than 20/70 vision and it includes the categories of partially sighted and blind; Partially sighted is defined as having visual acuity better than 20/200 up to and including 20/70 (Lowenfeld, 1973); Blind is defined as having less than 20/200 vision in the healthier eye (Bishop, 1996). No distinction was made between congenital or adventitious acquired visual impairment, or between blindness and partially sightedness.

Method

Searching and screening studies. Starting point of this meta-analysis was the dataset of te Nijenhuis, van Rijk, and Kämper (2007), who performed a meta-analysis testing the cumulative deficit theory. Eight methods were used to obtain IQ scores of visually impaired children. First, a manual article by article search was carried out in a large number of journals, such as *The International Journal for the Education of the Blind*, *Journal of Visual Impairment and Blindness*, *British Journal of Visual Impairment*, *Education of Visually Handicapped*, *Sehgeschädigte*, *Zeitschrift für das Blinden- und Sehbehindertenbildungswesen*, and *Research Bulletin: American Foundation for the Blind*, from 1930-2007. Additionally, an electronic search of the journal *Visual Impairment Research* was conducted. Second, an electronic search for published articles using PsycINFO, PiCarta, Academic search premier, ERIC, MEDLINE, Web of science, Google Scholar, Scirus, ABI-Inform, Tesionline, and PubMed was conducted. The following keyword combinations were used to conduct the searches: any keyword that contains the word “blind*”, “visual* impair*”, “visually handicapped, congenital blind*, blind person*/subject*, legally blind*, and “sighted” in combination with any keyword that contains one of the following words; IQ, *g*, general mental ability, GMA, cognitive ability, general cognitive ability, intelligence, intelligence test, Wechsler, Stanford Binet, cognitive ability test (* is a truncation symbol to represent multiple spellings or endings; AND is a Boolean operator that combines search terms so that the search result contains all of the terms). Third, manuals of tests for the assessment of visually impaired individuals were checked searching for mean IQs. Fourth, several institutions for the blind, namely Bartimèus and Visio in the Netherlands, and Center of Learning for Visually Impaired and Blind [BBS–Nürnberg] in Germany were contacted and

asked for studies that reported the mean IQ scores of blind and partially sighted individuals. Fifth, libraries and test libraries of the Universiteit van Amsterdam, Vrije Universiteit Amsterdam, Universiteit Groningen, Universiteit Nijmegen, Universität Dortmund, Humboldt-Universität, and Universität Heidelberg were visited in order to search for relevant articles. Sixth, the reference lists of all currently included empirical studies were studied to identify any articles that may have been missed by earlier search methods. Seventh, a cited reference search was conducted to identify new studies referring to studies that were already obtained. Finally, several authors were contacted in order to obtain any additional articles or supplementary information. After the te Nijenhuis et al. (2007) MA was completed we received the datasets of Studeny (2008), comparing blind children in Austria and South Africa. We added the data points to the MA.

Specific criteria for inclusion. To be included in the meta-analysis three additional criteria had to be met. First, only empirical studies reporting IQ test scores of partially sighted or blind children or adults were included. Second, the mean subtest scores had to be lower than the mean scores of the standardization sample of the IQ test. Finally, studies in which subjects had additional handicaps known to influence mental performance were excluded. Application of these inclusion rules yielded five studies resulting in six correlations between *g* and score differences between a visually impaired group and an average group.

Computation of score differences between a visually impaired group and an average group. Score differences between a visually impaired group and an average group (*d*) were computed by subtracting the mean score of the visually impaired group of the particular test in question from the mean score of the standardization group, and then dividing the result by the *SD* of the standardization group. The standardization group scores were obtained by computing a weighted average score matching the age range of the visually impaired group as closely as possible. The *g* loadings were obtained in the same way. The weighted average *g* loadings were computed, matching the age range of the visually impaired group to the age range of the *g* loadings as close as possible.

Psychometric meta-analytical techniques (Hunter & Schmidt, 1990, 2004) were applied to the resulting five *r* (*g* × visual impairment)s using the software package developed by Schmidt and Le (2004). In the present study we corrected for the five artifacts (mentioned above) that alter the value of outcome measures listed by Hunter and Schmidt (1990).

Correction for reliability of the vector of visual impairment. The value of *r* (*g* × visual impairment) is attenuated by the reliability of the vector of visual impairment for a given battery. The reliability of the vector of visual impairment was estimated using the present datasets,

comparing samples that took the same test and that were comparable in regard to age and sample size. As an illustration of the procedure, the following rules were set in order to analyze studies that were highly comparable. First of all, only studies using the same test and the same version of this test were taken together. Second, studies containing less than a hundred participants were considered to be highly comparable as long as the difference in N between two studies was less than or equal to sixty. Third, studies containing more than a hundred participants were considered to be highly comparable as long as the difference in N between two studies was lesser than or equal than hundred-fifty. Fourth, the difference in average age of participants in separate studies was three years or less. Finally, the date of publication between two studies did not differ more than ten years. Baitinger and Bernd (1970) and Rath (1967) reported IQ scores on six HAWIK subtests, which were used to construct the distribution of reliability of the vector of visual impairment, but which were not used as data points for the meta-analysis.

A scatter plot of reliabilities against Ns should reveal that the larger N becomes, the higher the value of the reliability coefficients, with an asymptotic function between r (d x d) and N expected. We checked to see which curve gave the best fit to the expected asymptotic function. Figure 4 shows the scatter plot of reliability of the vector of visual impairment and sample size, and the logarithmic curve that fitted optimally. Because of the small range in N the logarithmic regression line is almost linear.

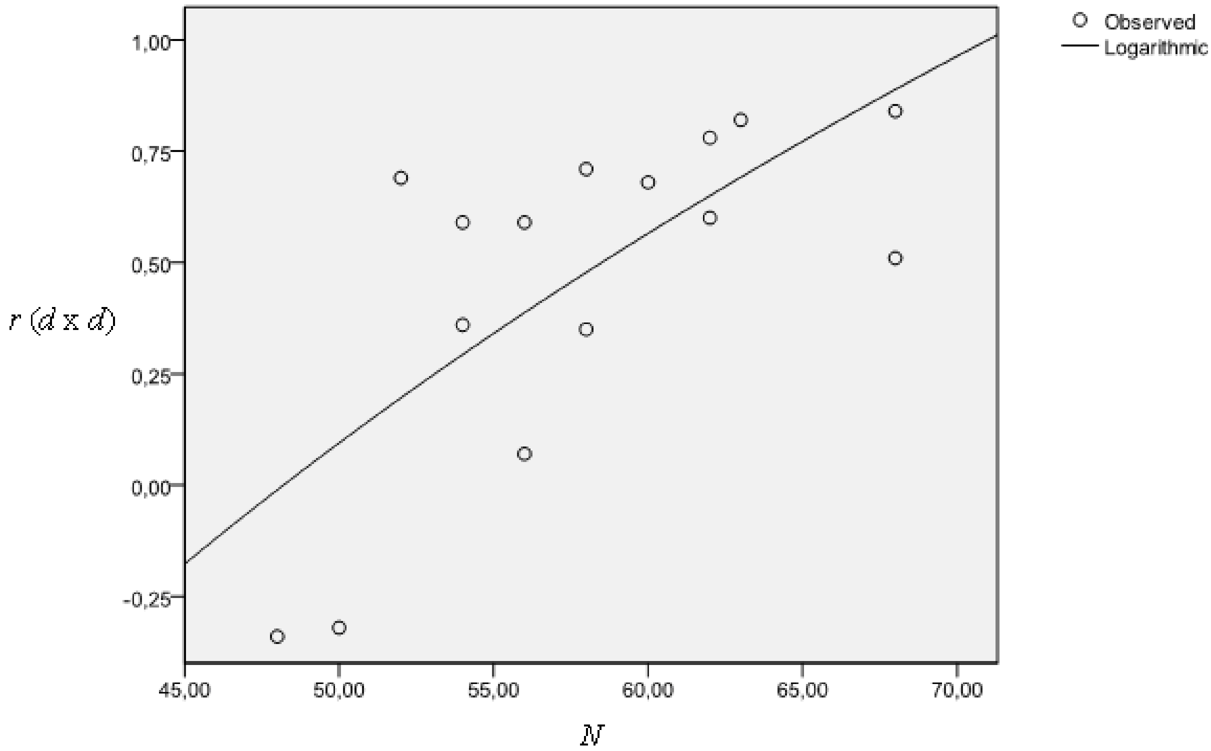

Figure 4

Scatter Plot of Reliability of the Vector of Visual Impairment and Sample Size and Regression Line

Results

The results of the studies on the correlation between *g* loadings and the score differences between visually impaired groups and average groups (*d*) are shown in Table 4. The Table gives data derived from five studies, with participants numbering a total of 363 predominantly young children. It also lists the reference for the study, the cognitive ability test used, the correlation between *g* loadings and *d*, the sample size, and the mean age (and range of age).

Table 5 presents the results of the psychometric meta-analysis of the six data points for young children. It shows (from left to right): the number of correlation coefficients (*K*), total sample size (*N*), the mean observed correlations (*r*) and their standard deviation (SD_r), the correlations one can expect once artifactual error from unreliability in the *g* vector, the visual impairment vector, and range restriction in the *g* vector have been removed (rho-4), and their standard deviation ($SD_{\text{rho-4}}$), and the true correlation one can expect when corrections for all five artifacts have been carried out (rho-5). The next two columns present the percentage of variance explained by artifactual errors (%VE), and the 80% confidence interval (80% CI). This interval

denotes the values one can expect for rho in sixteen out of twenty cases.

For the group of predominantly young children the analysis of all six data points yields an estimated correlation (rho-4) of .04, with only 5% of the variance in the observed correlations explained by artifactual errors. However, Hunter and Schmidt (1990) state that extreme outliers should be left out of the analysis, because they are most likely the result of errors in the data. They also argue that strong outliers artificially inflate the *SD* of effect sizes and thereby reduce the amount of variance that artifacts can explain. Figure 5 shows the scatter plot of correlations *r* (*d* x *g*) against sample size. We chose to leave out two extreme outliers, namely both groups by Studeny (2008), with *r* values of 5.5 *SD* and 6.2 *SD* respectively, for the South African and the Austrian sample, above the average *r* of the final sample of four data points. This resulted in a value of the correlation (rho-4) of -.65, a large decrease in the *SD* of rho-4 with 81%, and nearly a twelvefold increase of the amount of variance explained in rho-4 by artifacts: 64% of the variance is now explained. Finally, a correction for deviation from perfect construct validity in *g* took place, using the conservative value of .90. This resulted in a value of -.72 for the final estimated true correlation between *g* loadings and visual impairment in predominantly young children.

Table 4

Studies of Correlations Between g Loadings and Visual Impairment

reference	test	r ($d \times g$)	N	age mean (range)
Children mean age ≤ 13 years (-1.00 expected)				
Baitinger & Bernd (1970)	HAWIK	-.43	73	10.55[a] (6.0-15.1)
Daugherty & Moran (1982)	WISC-R	-.71	50	12.5[a] (7.0-18.0)
Klauer (1962)[b]	HAWIK	-.36	62	10.8[c] (6.1-15.6)[c]
Krüger (1974)	HAWIK	-.19	53	11.1[a] (6.1-16.1)
Studeny (2008)	HAWIK-IV[d]	.93	51	11.6[a] (10.0–13.25)
	WISC-IV[e]	.79	74	

Note. [a]Estimated value. [b]Reported in Baitinger & Bernd (1970).[c]Krüger (1974) combined his study with Klauer (1962), and Baitinger & Bernd (1970), from which age mean and range were estimated. [d]Austrian sample. [e]South African sample.

Table 5

Meta-analytical Results for the Correlation Between Visually Impairment and g Loadings After Corrections for Reliability, Restriction of Range, and Imperfect Construct Validity

predictor	*K*	*N*	*r*	SD_r	rho-4	$SD_{\text{rho-4}}$	rho-5	% VE	80% CI
mean age ≤ 13 years									
Visual impairment[1]	6	363	.01	.59	.04	.80	.04	5%	-1.01- 1.03
Visual impairment minus 2 outliers[2]	4	238	-.42	.14	-.65	.15	-.72	64%	-.84 - -.47

Note. [1]Meta-analytical results for correlations between *g* loadings and *d* (visual impairment). [2] The study by Studeny (2008) is considered an extreme outlier and therefore left out of the analysis. *K* = number of correlations; *N* = total sample size; *r* = mean observed correlation (sample-size weighted); SD_r = standard deviation of observed correlation; rho-4 = observed correlation corrected for sampling error, unreliability, and range restriction); $SD_{\text{rho-4}}$ = standard deviation of correlation; rho-5= true correlation (observed correlation corrected for sampling error, unreliability, range restriction, and imperfect construct validity); %VE = percentage of variance accounted for by artifactual errors; 80% CI = 80% credibility interval.

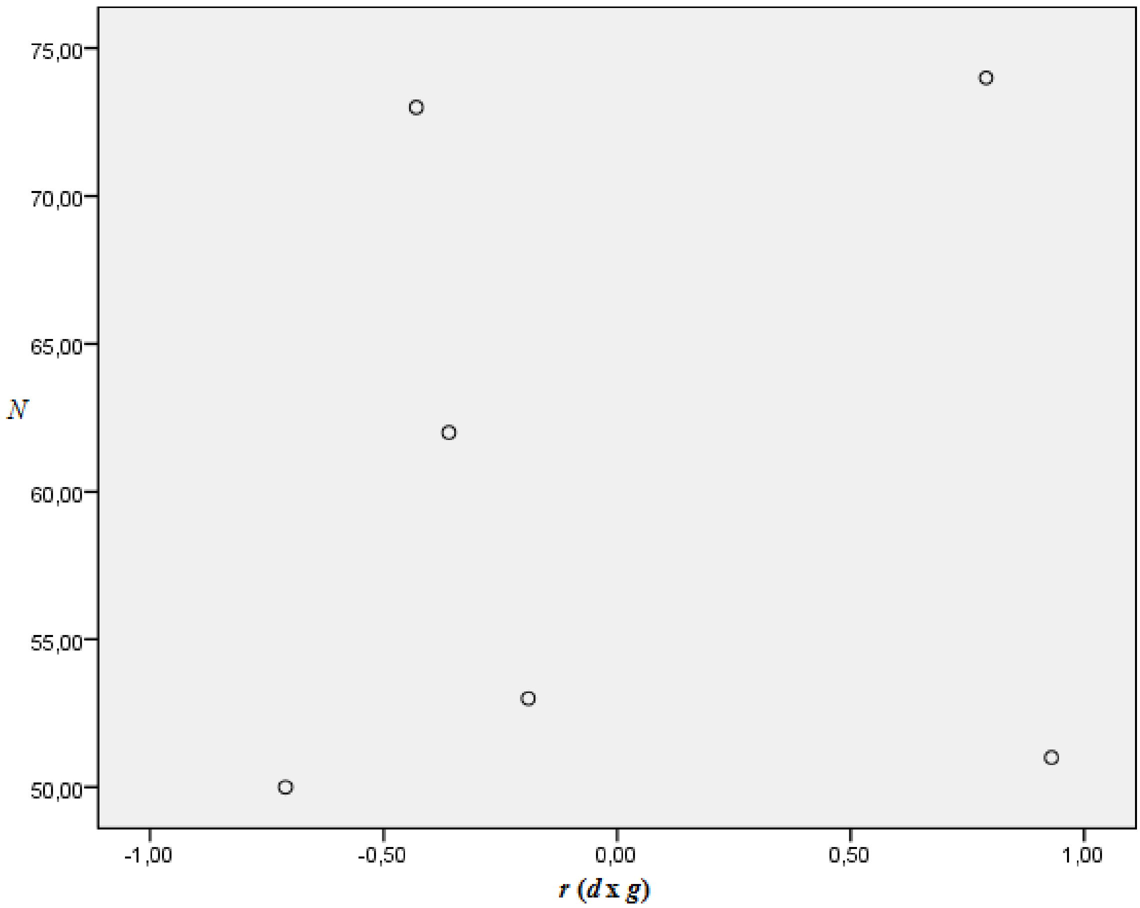

Figure 5

Scatterplot of Correlations r (d x g) and Sample Size of the Variable Visual Impairment

Conclusion

It is concluded that the hypothesis is supported quite strongly. Four of the six datasets show the expected negative correlation. Leaving out two outliers results in a rho-5 of -.72, while a theoretical value of -1.00 was expected for samples existing only of young children. However, the present samples also contained substantial proportions of older children. This could lead to a correlation slightly above -1.00, and that's exactly what we found here.

Study 3: Hearing impairment

To test whether there is a –1.00 correlation between the magnitude of the *g* loadings of IQ subtests and the scores of hearing-impaired children with an average age ≤ 13.00 years an exploratory bare-bone psychometric meta-analysis was performed. A bare-bone psychometric meta-analysis estimates how much of the observed variance in findings across studies is due to sample size alone. Hearing impairment is defined as having a hearing loss of at least 70 dB (pure tone average) in the better ear (Anderson & Sisco, 1977). All subjects were congenitally or prelingually deaf.

Method

Searching and screening studies. The studies for this exploratory meta-analysis are derived from Braden (1989; Table 1, pp. 151).

Specific criteria for inclusion. For a study to be included in the current meta-analysis the following additional criterion had to be met: subtest scores following hearing impairment had to be lower than the average subtest scores of a hearing comparison group. Based on this criterion two studies were excluded. Two studies were suited for inclusion in the exploratory meta-analysis.

Computation of score differences between a hearing impaired group and a hearing group. Score differences between a hearing impaired group and a hearing group (*d*) were computed by subtracting the mean of the hearing impaired group from the mean of the hearing group, and then dividing the result by the (mean) *SD* of the standardization group(s) of the particular test in question.

Bare-bones meta-analytical techniques (Hunter & Schmidt, 1990, 2004) were applied to the resulting 2 r ($g \times d$)s using the software package developed by Schmidt and Le (2004).

Results

The results of the studies on the correlation between *g* loadings and the score differences between hearing impaired groups and hearing groups (*d*) are shown in Table 6. The Table gives

data derived from two studies, with participants numbering a total of 1,287. It also lists the reference for the study, the cognitive ability test used, the correlation between *g* loadings and *d*, and the sample size. It is clear that both correlations are quite strongly negative in sign. Table 7 presents the results of the bare-bones MA of the two data points. It shows (from left to right): the number of correlation coefficients (*K*), total sample size (*N*), the true correlation (rho) and their standard deviation (SD_r) The last column present the percentage of variance explained by artifactual errors (%VE). The analysis of both data points yields an estimated correlation (rho) of -.69, with 277% of the variance in the observed correlations explained by artifactual errors.

This phenomenon is called "second-order sampling error", and results from the sampling of studies in a meta-analysis. Percentages of variance explained greater than 100% are not uncommon when only a limited number of studies are included in an analysis. The proper conclusion is that all the variance is explained by statistical artifacts (see Hunter & Schmidt, 2004, pp. 399-401, for an extensive discussion).

Table 6
Studies of Correlations Between g Loadings and Hearing impairment

reference	test	*r*	*N*
Braden (1984)[a]	WISC-R	-.69	59
Hirshoren, Hurley, & Kavale (1979)	WISC-R	-.63	1228

Note. [a] Data derived from Anderson and Sisco (1977).

Table 7
Exploratory Bare-bones Meta-analytical Results for Correlations Between g Loadings and Deaf-Hearing Score Differences

predictor	*K*	*N*	rho	SD_{rho}	%VE
deaf-hearing[1]	2	1,287	-.69	.00	277%

Note. [1]Bare-bones meta-analytical results: Score differences between a deaf group, a hearing group, and *g* loadings. *K* = number of correlations; *N* = total sample size; rho = true correlation (observed correlation corrected for sample size); SD_{rho} = standard deviation of true correlation; %VE = percentage of variance accounted for by artifactual errors.

Conclusion

Like visual impairment, hearing impairment in children ≤ 13.00 years goes with lower IQ scores. We explored whether this lower IQ was related to *g*, and both datasets show the expected negative correlation. A true correlation of -.69 was found, while a theoretical value of -1.00 was expected for samples existing only of young children. However, the present samples also contained substantial proportions of older children. This could lead to a correlation slightly below -1.00, and that's exactly what we found here.

Study 4: Schizophrenia

To test whether there is a correlation between the magnitude of *g* loadings and IQ scores of schizophrenics, an exploratory psychometric meta-analysis was performed on a number of studies that reported IQ scores of at least seven subtests from schizophrenic subjects.

Method

Searching and screening studies. Three methods were used to identify studies that contained IQ scores of schizophrenics. First, an electronic search for published research using PsycINFO, ERIC, MEDLINE, PiCarta, Academic search premier, Web of science, Google Scholar, and PubMed was conducted. Keywords used were schizophren*, and, cognitive, mental*, intelligence, IQ, WISC, Wechsler, and combinations of these concepts (* is a truncation symbol to represent multiple spellings or endings; AND is a Boolean operator that combines search terms so that the search result contains all of the terms). Second, the reference lists of significant articles were analyzed in search of additional studies. Finally, cited reference searches were conducted using Web of Science, to search for articles citing significant articles. This procedure resulted in thirteen articles, book chapters, and reports on the concurrent topics of schizophrenia and mental ability. Five studies met all criteria for inclusion in the meta-analysis.

Specific criteria for inclusion. For a study to be included in the meta-analysis, three additional criteria had to be met: First, only empirical studies reporting IQ test scores on schizophrenics were included. Second, the mean subtest scores had to be lower than the mean scores of the standardization sample of the IQ test. Finally, studies that reported on comorbid disorders known to influence mental performance, such as ADHD were excluded. The subjects from the study by Nelson, Pantelis, Carruthers, Speller, Baxendale, and Barnes (1990) were inpatients in a mental hospital, but no comorbid disorders are reported.

Computation of score differences between a schizophrenic group and an average group. Score differences between a schizophrenic group and an average group (*d*) were computed by subtracting the mean score of the schizophrenic group of the particular test in question from the mean score of the standardization group, and then dividing the result by the *SD* of the standardization group. The standardization group scores were obtained by computing a weighted average score matching the age range of the schizophrenic group as closely as possible. The *g* loadings were obtained in the same way. The weighted average *g* loadings were computed, matching the age range of the schizophrenic group to the age range of the *g* loadings as close as possible.

In the study by Bilder, Lipschutz-Broch, Reiter, Geisler, Mayerhoff, and Lieberman (1992) we decided to compute a weighted average of a group of first-episode schizophrenics (n=51), and a group of chronic schizophrenics (n=50). Hunter & Schmidt (2004) advice to collate data to increase sample size when studies contain several small sample size groups, if this doesn't dramatically alter the outcome. Goldberg, Ragland, Torrey, Gold, Bigelow, and Weinberger (1990) compared sixteen schizophrenic patients with their monozygotic twin siblings (total N=32). Since monozygotic twins are genetically identical, and therefore provide by far the best comparison groups to investigate genetic variables, we choose to use the twin siblings to compute d values. The sample size (n=16) was multiplied by a factor five which increases the weight of this study within the meta-analysis.

Psychometric meta-analytical techniques (Hunter & Schmidt, 1990, 2004) were applied to the resulting five (g × schizophrenia)s using the software package developed by Schmidt and Le (2004). In the present study we corrected for the five artifacts (mentioned above) that alter the value of outcome measures listed by Hunter and Schmidt (1990).

Correction for the reliability of the vector of schizophrenia. The value of r (g × schizophrenia) is attenuated by the reliability of the vector of schizophrenia for a given battery. The reliability of the vector of schizophrenia was estimated using the present datasets, comparing samples that took the same test and that were comparable in regard to age and sample size. As an illustration of the procedure, the following rules were set in order to analyze studies that were highly comparable. First of all, only studies using the same test and the same version of this test were taken together. Second, studies containing less than a hundred participants were considered to be highly comparable as long as the difference in N between two studies was lesser than or equal to sixty. Third, studies containing more than a hundred participants were considered to be highly comparable as long as the difference in N between two studies was lesser than or equal than hundred-fifty. Fourth, the difference in average age of participants in separate studies was three years or less. Finally, the date of publication between two studies did not differ more than ten years.

A scatter plot of reliabilities against Ns should reveal that the larger N becomes, the higher the value of the reliability coefficients, with an asymptotic function between r (d x d) and N expected. We checked to see which curve gave the best fit to the expected asymptotic function. Figure 6 shows the scatter plot of reliability of the vector of schizophrenia and sample size, and the logarithmic curve that fitted optimally.

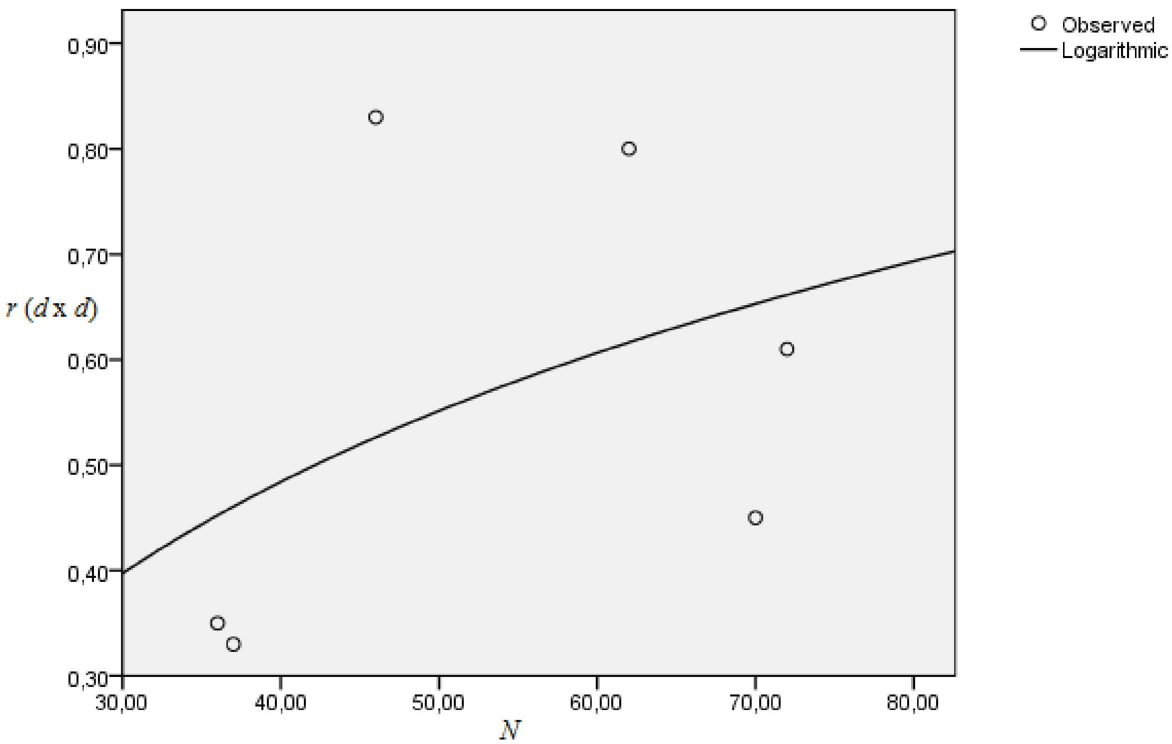

Figure 6

Scatter Plot of Reliability of the Vector of the Schizophrenia and Sample Size and Regression Line

Results

The results of the studies on the correlation between *g* loadings and the score differences between schizophrenic groups and average groups (*d*) are shown in Table 8. The Table gives data derived from five studies, with participants numbering a total of 315. It also lists the reference for the study, the cognitive ability test used, the correlation between *g* loadings and *d*, the sample size, and the mean age (and range of age). It is clear that the majority of the correlations are quite strongly negative in sign.

Table 9 presents the results of the psychometric meta-analysis of the four data points. It shows (from left to right): the number of correlation coefficients (*K*), total sample size (*N*), the mean observed correlations (*r*) and their standard deviation (SD_r), the correlations one can expect once artifactual error from unreliability in the *g* vector, the schizophrenia vector, and range restriction in the *g* vector have been removed (rho-4), and their standard deviation ($SD_{\text{rho-4}}$), and the true correlation one can expect when corrections for all five artifacts have been carried out (rho-5). The next two columns present the percentage of variance explained by artifactual errors (%VE), and the 80% confidence interval (80% CI). This interval denotes the values one can

expect for rho in sixteen out of twenty cases.

The analysis of all five data points yields an estimated correlation (rho-4) of -.78, with 138% of the variance in the observed correlations explained by artifactual errors. This phenomenon is called "second-order sampling error", and results from the sampling of studies in a meta-analysis. Percentages of variance explained greater than 100% are not uncommon when only a limited number of studies are included in an analysis. The proper conclusion is that all the variance is explained by statistical artifacts (see Hunter & Schmidt, 2004, pp. 399-401, for an extensive discussion).

Finally, a correction for deviation from perfect construct validity in *g* took place, using the conservative value of .90. This resulted in a value of -.87 for the final estimated true correlation between *g* loadings and schizophrenia.

Table 8

Studies of Correlations Between g Loadings and Schizophrenia

reference	test	*N*	*r* ($d \times g$)	age mean (range)
Bilder et al. (1992)	WAIS-R	101	-.48	28.7[a]
Di Nuovo & Buono (2007)	WAIS-R	11	-.43	40.5[b] (17-64)
Goldberg et al. (1990)[c]	WAIS-R / WMS-R	80	-.65	31.5 (19-44)
Morice (1990)	WAIS-R	60	-.32	32.0[a]
Nelson et al. (1990)	WAIS-R	63	-.53	50.2[a]

Note. [a]Age range not reported. [b]Estimated value. [c]The *r* value was based upon comparison to the control group (monozygotic twins; *n*= 16). The sample size in this study (*n*=16) was multiplied by 5, because there was a perfect matched control group (*n*=16), namely monozygotic twins).

Table 9

Meta-analytical Results for the Correlation Between Schizophrenia and g Loadings after Corrections for Reliability, Restriction of Range, and Imperfect Construct Validity

Predictor	*K*	*N*	r	SD_r	rho-4	$SD_{\text{rho-4}}$	rho-5	% VE	80% CI
Schizophrenia[1]	5	315	-.50	.06	-.78	.00	-.87	138%	-.78 - -.78

Note. [1]Meta-analytical results for correlations between *g* loadings and *d* (schizophrenia). *K* = number of correlations; *N* = total sample size; *r* = mean observed correlation (sample-size weighted); SD_r = standard deviation of observed correlation; rho-4 = observed correlation corrected for sampling error, unreliability, and range restriction); $SD_{\text{rho-4}}$ = standard deviation of correlation; rho-5= true correlation (observed correlation corrected for sampling error, unreliability, range restriction, and imperfect construct validity); %VE = percentage of variance accounted for by artifactual errors; 80% CI = 80% credibility interval.

Conclusion

Schizophrenia goes with lower IQ. We explored whether this lower IQ was related to *g*, and contradicting our hypothesis, found a negative correlation. However, after a detailed study of the five data points it is clear that the scores on Performal subtests are lower than the scores on Verbal subtests. Performal subtests on average have lower *g* loadings than Verbal subtests, which strongly influences the correlation. We conclude that there is a methodological weakness in the MCV: a strong negative correlation can result from both (1) a clear, linear negative relation between *g* and *d*, and (2) stronger effects on the Performal subtests than the Verbal subtests.

Study 5: Epilepsy

To test whether there is a correlation between the magnitude of *g* loadings and IQ scores of epileptics, an exploratory psychometric meta-analysis was performed on a number of studies that reported IQ scores of at least seven subtests from epileptic subjects.

Method

Searching and screening studies. Three methods were used to identify studies that contained IQ scores of epileptics. First, an electronic search for published research using PsycINFO, ERIC, MEDLINE, PiCarta, Academic search premier, Web of science, Google Scholar, and PubMed was conducted. Keywords used were epilep*, and, cognitive, mental*, intelligence, IQ, WISC, Wechsler, and combinations of these concepts (* is a truncation symbol to represent multiple spellings or endings; AND is a Boolean operator that combines search terms so that the search result contains all of the terms). Second, the reference lists of significant articles were analyzed in search of additional studies. Finally, cited reference searches were conducted using Web of Science, to search for articles citing significant articles. This procedure resulted in fifteen articles, book chapters, and reports on the concurrent topics of epilepsy and mental ability. Seven studies met all criteria for inclusion in the meta-analysis.

Specific criteria for inclusion. For a study to be included in the meta-analysis, three additional criteria had to be met: First, only empirical studies reporting IQ test scores on epileptics were included. Second, the mean subtest scores had to be lower than the mean scores of the standardization sample of the IQ test. Finally, studies that reported on comorbid conditions known to influence mental performance, such as ADHD or learning disability were excluded.

Computation of score differences between an epileptic group and an average group. Score differences between an epileptic group and an average group (*d*) were computed by subtracting the mean score of the epileptic group of the particular test in question from the mean

score of the standardization group, and then dividing the result by the *SD* of the standardization group. The standardization group scores were obtained by computing a weighted average score matching the age range of the epileptic group as closely as possible. The *g* loadings were obtained in the same way. The weighted average *g* loadings were computed, matching the age range of the epileptic group to the age range of the *g* loadings as close as possible.

Psychometric meta-analytical techniques (Hunter & Schmidt, 1990, 2004) were applied to the resulting seven (*g* × epilepsy)s using the software package developed by Schmidt and Le (2004). In the present study we corrected for the five artifacts (mentioned above) that alter the value of outcome measures listed by Hunter and Schmidt (1990).

Correction for the reliability of the vector of epilepsy. The value of *r* (*g x* epilepsy) is attenuated by the reliability of the vector of epilepsy for a given battery. The reliability of the vector of schizophrenia was estimated using the present datasets, comparing samples that took the same test and that were comparable in regard to age and sample size. As an illustration of the procedure, the following rules were set in order to analyze studies that were highly comparable. First of all, only studies using the same test and the same version of this test were taken together. Second, studies containing less than a hundred participants were considered to be highly comparable as long as the difference in *N* between two studies was lesser than or equal to sixty. Third, studies containing more than a hundred participants were considered to be highly comparable as long as the difference in *N* between two studies was lesser than or equal than hundred-fifty. Fourth, the difference in average age of participants in separate studies was three years or less. Finally, the date of publication between two studies did not differ more than ten years.

A scatter plot of reliabilities against *N*s should reveal that the larger *N* becomes, the higher the value of the reliability coefficients, with an asymptotic function between *r* (*d* x *d*) and *N* expected. We checked to see which curve gave the best fit to the expected asymptotic function. Figure 7 shows the scatter plot of reliability of the vector of epilepsy and sample size, and the logarithmic curve that fitted optimally.

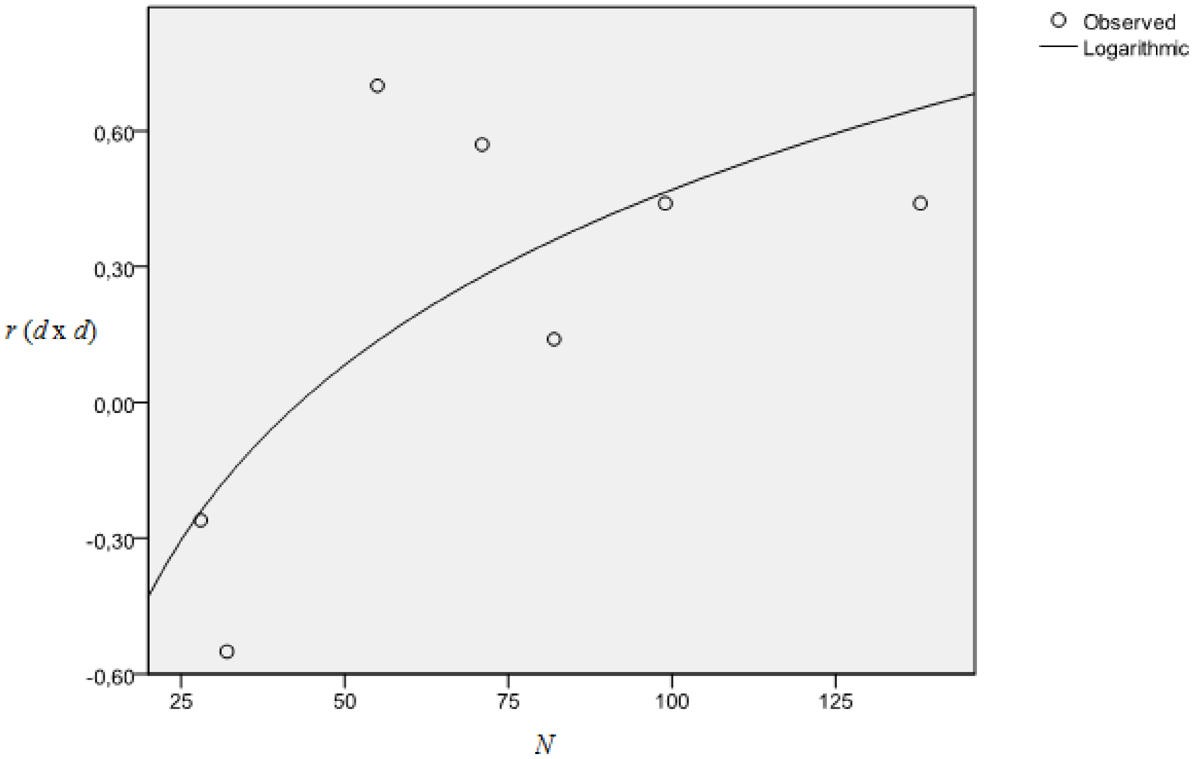

Figure 7

Scatter Plot of Reliability of the Vector of the Epilepsy and Sample Size and Regression Line

Table 10

Studies of Correlations Between g Loadings and Epilepsy

reference	Test	*r* ($d \times g$)	*N*	*age mean (range)*
Adachi et al. (2005)	WAIS-R	.70	55	35.0
Baker, Austin, & Downes (2003)	WMS-III	.44	99	34
Barr (1997)	WMS-R	.14	82	33.1
Kim, Yi, Son & Kim (2003)	K-WAIS	.57	71	29.3 (22-36)
Moore & Baker (1996)	WMS-R	.44	138	31.6
O'Leary, Burns & Borden (2006)	WISC-III	-.55	32	11.0[a] (6-16)
Schneider, Nowack, Fitzgerald, Janati & Souheaver (1993)	WAIS	-.26	28	46.1

Note. [a]Estimated value.

Table 11

Meta-analytical Results for the Correlation Between Epilepsy and g Loadings after Corrections for Reliability, Restriction of Range, and Imperfect Construct Validity

predictor	*K*	*N*	*R*	SD_r	rho-4	$SD_{\text{rho-4}}$	rho-5	% VE	80% CI
Epilepsy[1]	7	505	.34	.30	.46	.36	.51	28%	.00-.92
minus one outlier[2]	6	473	.37	.28	.51	.31	.57	36%	.11-.90
minus two outiers[3]	5	445	.44	.14	.58	.00	.64	137%	.58-.58

Note. [1]Meta-analytical results for correlations between *g* loadings and *d* (epilepsy). *K* = number of correlations; *N* = total sample size; *r* = mean observed correlation (sample-size weighted); SD_r = standard deviation of observed correlation; rho-4 = observed correlation corrected for sampling error, unreliability, and range restriction); $SD_{\text{rho-4}}$ = standard deviation of correlation; rho-5= true correlation (observed correlation corrected for sampling error, unreliability, range restriction, and imperfect construct validity); %VE = percentage of variance accounted for by artifactual errors; 80% CI = 80% credibility interval.

Results

The results of the studies on the correlation between *g* loadings and the score differences between epileptic groups and average groups (*d*) are shown in Table 10. The Table gives data derived from seven studies, with participants numbering a total of 505. It also lists the reference for the study, the cognitive ability test used, the correlation between *g* loadings and *d*, the sample size, and the mean age (and range of age). It is clear that the majority of the correlations are quite strongly negative in sign.

Table 11 presents the results of the psychometric meta-analysis of the seven data points. It shows (from left to right): the number of correlation coefficients (*K*), total sample size (*N*), the mean observed correlations (*r*) and their standard deviation (SD_r), the correlations one can expect once artifactual error from unreliability in the *g* vector, the epilepsy vector, and range restriction in the *g* vector have been removed (rho-4), and their standard deviation ($SD_{\text{rho-4}}$), and the true correlation one can expect when corrections for all five artifacts have been carried out (rho-5). The next two columns present the percentage of variance explained by artifactual errors (%VE), and the 80% confidence interval (80% CI). This interval denotes the values one can expect for rho in sixteen out of twenty cases.

The analysis of all seven data points yields an estimated correlation (rho-4) of .46, with 28% of the variance in the observed correlations explained by artifactual errors. However, Hunter and Schmidt (1990) state that extreme outliers should be left out of the analysis, because they are most likely the result of errors in the data. They also argue that strong outliers artificially inflate the *SD* of effect sizes and thereby reduce the amount of variance that artifacts can explain. Figure 8 shows the scatter plot of correlations *r* (*d* x *g*) against sample size. We chose to leave out two extreme outliers, namely the studies by O'Leary, Burns and Borden (2006) and Schneider, Nowack, Fitzgerald, Janati, and Souheaver (1993) with *r* values of 4.8 *SD* and 3.5 *SD,*

respectively, below the average *r* of the final sample of five data points. This resulted in a value of the correlation (rho-4) of .58, a large decrease in the *SD* of rho-4 with 36%, and nearly a fourfold increase of the amount of variance explained in rho-4 by artifacts: 137% of the variance is now explained.

Finally, a correction for deviation from perfect construct validity in *g* took place, using the conservative value of .90. This resulted in a value of .64 for the final estimated true correlation between *g* loadings and epilepsy.

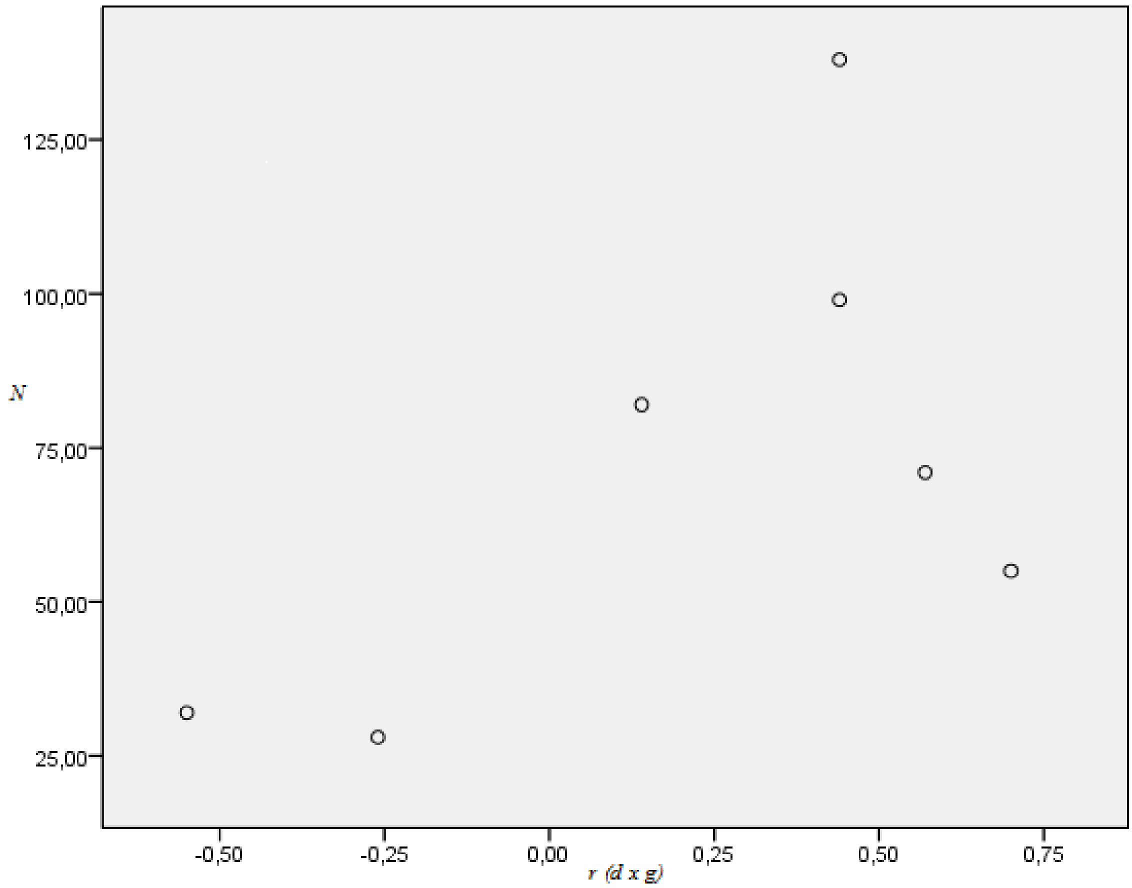

Figure 8

Scatterplot of Correlations r (d x g) and Sample Size of the Variable Epilepsy

Conclusion

The exploratory MA shows that the effects of epilepsy are quite strongly linked to the *g* vector. Table 11 shows a percentage of variance explained of 137%.This phenomenon is called “second-order sampling error”, and results from the sampling of studies in a meta-analysis. Percentages of variance explained greater than 100% are not uncommon when only a limited

number of studies are included in an analysis. The proper conclusion is that all the variance is explained by statistical artifacts (see Hunter & Schmidt, 2004, pp. 399-401, for an extensive discussion).

Study 6: A methodological experiment using data on Giftedness

The sixth MA consisted of a methodological experiment. The data from the psychometric MA on giftedness by te Nijenhuis, de Pater, van Bloois, and Geutjes (2009) were reanalyzed, this time using all Wechsler composite *Verbal*, *Performance*, and *Full Scale* IQ scores to compute *d* values and *g* loadings. Using part of the same dataset enabled us to investigate the robustness of the MCV, when *V*, *P*, and *FS* IQ scores are used, instead of a minimum of seven subtests. Wechsler composite *V*, *P*, and *FS* IQ scores are frequently reported in published articles, while scores on subtest level are rare.

Method

Searching and screening studies. The present MA was based entirely on the dataset put at our disposal by te Nijenhuis, et al. (2009). We concluded that the search had been thorough, so no additional studies were sought.

Both electronic and manual searches were conducted by te Nijenhuis et al. for studies that contained cognitive ability data of the gifted. Four methods were used to obtain scores of the gifted from published studies for the present meta-analysis. First, an electronic search for published research using PsycINFO, PiCarta, Academic search premier, Web of science, and PubMed was conducted. The following combinations were used to conduct the searches: any keyword that contains the word “gifted”, or “exceptional” in combination with any keyword that contains one of the following words; IQ, *g*, general mental ability, GMA, cognitive ability, general cognitive ability, intelligence, intelligence test, Wechsler, Stanford Binet, cognitive ability test. Second, they browsed the tables of content of several major research journals with a strong focus on the gifted: *Psychology in the Schools* 1964-2008, *Gifted Child Quarterly* 1977-2008, *Roeper Review* 1990-2008, *Journal for Advanced Academics* 1996-2008, and *Exceptional Children* 1934-2008. Third, they checked the reference list of all currently included empirical studies to identify any potential articles that may have been missed by earlier search methods. Finally, several well-known researchers who have conducted cognitive ability research of the gifted were contacted in order to obtain any additional articles or supplementary information.

Specific criteria for inclusion. Two criteria had to be met for inclusion in the current MA. First, only WISC-R studies were selected. Second, only studies reporting a mean Full Scale IQ

score of 125 or higher were included in the meta-analysis. Application of the inclusion rules yielded sixteen datasets resulting in sixteen correlations between *g* and score differences between a gifted group and an average group. Te Nijenhuis et al. (2009) used 22 studies, so we used 73 % of their dataset.

Computation of score differences between a gifted group and an average group. All articles reported composite Verbal, Performance, and Full Scale IQ scores. Score differences between a gifted group and an average group (*d*) were computed by subtracting a fixed mean value of *M*=100.0 from the particular mean composite Verbal, Performance, and Full Scale IQ score of the gifted group in question, and then dividing the result by a fixed standard deviation value of *SD*=15.0. *g* Loadings were matched as closely as possible using both the average age and age range of the gifted group. The matched weighted average g loadings were then converted into composite Verbal, Performance, and Full Scale IQ *g* loadings, using Jensen's (1998, pp. 103/104) formula:

$$\{1+(\sum[r^2_{sg}/(1-r^2_{sg})]^{-1})\}^{-0.5}$$

where

r^2_{sg} = each subtest's squared *g* loading.

Psychometric meta-analytical techniques (Hunter & Schmidt, 1990, 2004) were applied to the resulting sixteen *r* (*g* × giftedness)s using the software package developed by Schmidt and Le (2004). In the present study we corrected for the five artifacts (mentioned above) that alter the value of outcome measures listed by Hunter and Schmidt (1990).

Correction for Reliability of the Vector of g Loadings. The value of r (g × giftedness) is attenuated by the reliability of the vector of g loadings for a given battery. When two samples have a comparable N, the average correlation between vectors is an estimate of the reliability of each vector. Several samples that differed little on background variables were compared. We chose samples that were highly comparable with regard to age. Samples of children in the age of 3 to 5 years were compared against other samples of children who did not differ more than 0.5 year of age. Samples of children in the age of 6 to 17 years were compared against other samples of children who did not differ more than 1.5 year of age.

Correlation matrices were collected from test manuals, books, articles, and technical reports. The dataset contains data on the WISC-R from the U.S., Canada, the Netherlands, Norway, Belgium, and Korea. This resulted in about 134 data points, which yielded 67

comparisons of g loadings of comparable groups from which to estimate the reliability for that group. To give an illustration of the procedure, van Haasen et al. (1986) report correlation matrices of the Dutch and the Flemish WISC-R for 22 samples in the age of 6-16 years. Samples of children in the age of 6 to 17 years were compared to other samples of children who do not differ by more than 1.5 years. Because the samples of children reported in van Haasen et al. (1986) were between 6 and 17 years only children were compared who did not differ more than 1.5 years. The *N*s in these samples were comparable. The resulting average correlation was .78 (combined $N = 3{,}018$; average $N = 137$).

A scatter plot of reliabilities against *N*s should show that the larger N becomes, the higher the value of the reliability coefficients, with an asymptotic function between r (g x g) and N expected. The curve that gave the best fit to the expected asymptotic function was selected. The logarithmic regression line resembled quite well the expected asymptotic distribution for reliabilities. Figure 9 shows the scatter plot of reliability of the vector of g loadings and sample size, and the logarithmic curve that fitted optimally.

Correction for reliability of the vector of giftedness. The value of r (g × giftedness) is attenuated by the reliability of the vector of giftedness for a given battery. It was estimated using the present datasets, comparing samples that took the same test and that were comparable in regard to age and sample size. As an illustration of the procedure, the following rules were set in order to analyze studies that were highly comparable. First of all, only studies using the same test and the same version of this test were taken together. Second, studies containing less than a hundred participants were considered to be highly comparable as long as the difference in N between two studies was lesser than or equal to sixty. Third, studies containing more than a hundred participants were considered to be highly comparable as long as the difference in N between two studies was lesser than or equal than hundred-fifty. Fourth, the difference in average age of participants in separate studies was three years or less. Finally, the date of publication between two studies did not differ more than ten years.

A scatter plot of reliabilities against *N*s should reveal that the larger N becomes, the higher the value of the reliability coefficients, with an asymptotic function between r (d x d) and N expected. We checked to see which curve gave the best fit to the expected asymptotic function. Figure 9 shows the scatter plot of reliability of the vector of giftedness and sample size, and the logarithmic curve that fitted optimally.

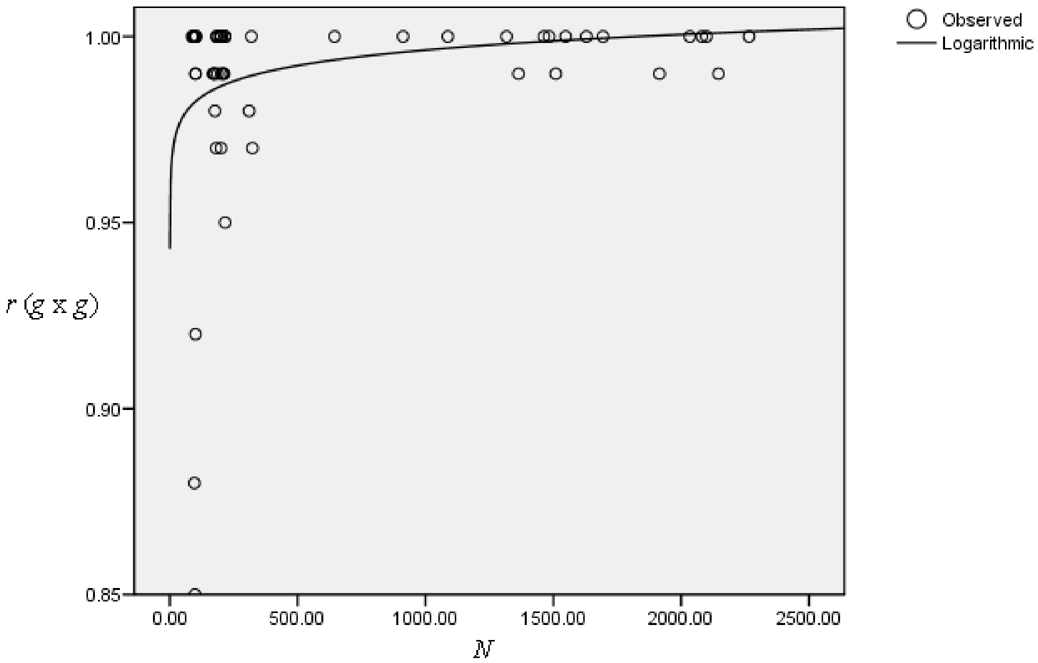

Figure 9

Scatter Plot of Reliability of the Vector of g Loadings and Sample Size and Regression Line

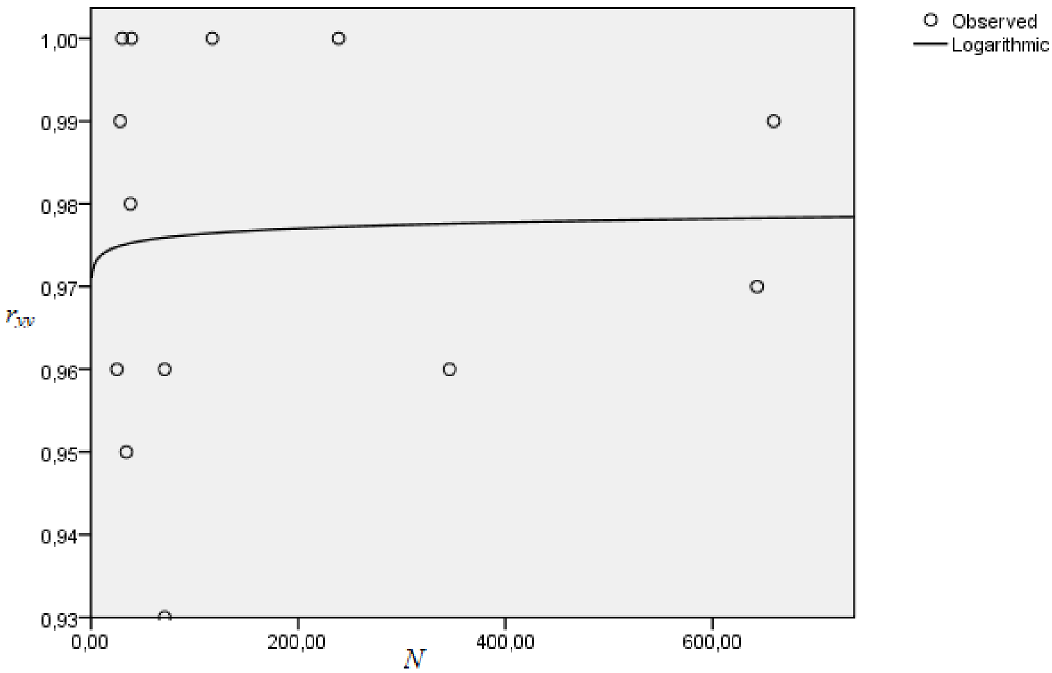

Figure 10

Scatter Plot of Reliability of the Vector of Giftedness and Sample Size and Regression Line

Results

The results of the studies on the correlation between *g* loadings and the score differences between gifted groups and average groups (*d*) are shown in Table 11. The Table gives data derived from sixteen studies, with participants numbering a total of 3,543. It also lists the reference for the study, the cognitive ability test used, the correlation between *g* loadings and *d*, the sample size, and the mean age (and range of age). It is clear that the majority of the correlations are strongly positive.

Table 12 presents the results of the psychometric meta-analysis of the sixteen data points. It shows (from left to right): the number of correlation coefficients (*K*), total sample size (*N*), the mean observed correlations (*r*) and their standard deviation (SD_r), the correlations one can expect once artifactual error from unreliability in the *g* vector, the giftedness vector, and range restriction in the *g* vector have been removed (rho-4), and their standard deviation ($SD_{\text{rho-4}}$), and the true correlation one can expect when corrections for all five artifacts have been carried out (rho-5). The next two columns present the percentage of variance explained by artifactual errors (%VE), and the 80% confidence interval (80% CI). This interval denotes the values one can expect for rho in sixteen out of twenty cases.

The analysis of all sixteen data points yields an estimated correlation (rho-4) of .93, with 35% of the variance in the observed correlations explained by artifactual errors. Finally, a correction for deviation from perfect construct validity in *g* took place, using the conservative value of .90. This resulted in a value of 1.03 for the final estimated true correlation between *g* loadings and giftedness.

Table 11
Gifted Studies using the WISC-R

study	*N*	age range	age *m*	*r*
Wechsler (1991)[b]	23	7.0-14.0	10.5[a]	,92
Wheaton & Vandergriff (1978)	26	9.5-11.5	10.8	.99
Ingram & Hakari (1985)	33	8.3-10.4	9.4[a]	.82
Sevier & Bain (1994)	35	7.0-12.7	9.1	,86
Henry & Wittman (1981)	40	6.0-12.0[a]	9.0[a]	,95
Phelps (1989)	48	7.7-15.8	11.6	1,00
Sabatino, Spangler, & Vance (1995)	51	9.2-16.3	13.6	,77
Reams, Chamrad, & Robinson (1990)	66	6.0-13.1	8.6	1,00
Spangler & Sabatino (1995)	66	6.0-12.0[a]	8.3	,84
Robinson & Nagle (1992)	75	7.0-14.0[a]	9.3	,98
Brown, Hwang, Baron, & Yakimowski (1991)	158	6.0-12.0[a]	9.6	,96
Brown & Yakimowksi (1987)	320	5.3-16.9	10.2	,97
Sapp, Chissom, & Graham (1985)	371	7.5-11.5	9.5	,87
Wilkinson (1993)	456	7.4-9.8	8.8	1,00
Macmann, Plasket, Barnett, & Siler (1991)	829	6.0-14.9	9.2	,79
Karnes & Brown (1980)	946	6.0-16.0	9.9	,94

Note. [a]Estimated value. [b]Reported in Sevier & Bain (1994).

Table 12
Meta-analytical Results for the Correlation Between Gifted and g Loadings After Corrections for Reliability, Restriction of Range, and Imperfect Construct Validity

predictor	*K*	*N*	*r*	SD_r	rho-4	$SD_{\text{rho-4}}$	rho-5	% VE	80% CI
Gifted	16	3.543	.91	.08	.93	.06	1.03	38%	.85-1.00

Note. [1]Meta-analytical results for correlations between *g* loadings and *d* (gifted). *K* = number of correlations; *N* = total sample size; *r* = mean observed correlation (sample-size weighted); SD_r = standard deviation of observed correlation; rho-4 = observed correlation corrected for sampling error, unreliability, and range restriction); $SD_{\text{rho-4}}$ = standard deviation of correlation; rho-5= true correlation (observed correlation corrected for sampling error, unreliability, range restriction, and imperfect construct validity); %VE = percentage of variance accounted for by artifactual errors; 80% CI = 80% credibility interval.

Conclusion

The validity of a simplified procedure for the MCV was tested on a dataset of gifted children, previously meta-analyzed using the MCV based on at least seven subtests (te Nijenhuis et al., 2009). This previous meta-analysis resulted in a rho-5 of 1.01, with 57 % of the variance in the datasets explained by five statistical artifacts. The simplified procedure for the MCV shows a value of rho-5 which is almost identical to the one previously found.

The simplified procedure yields only 38 % variance explained in the datasets. The more extensive procedure shows a value of 57 %, which is much higher than the value of 38 %. However, the simplified procedure was applied to only a subset of the larger dataset. As meta-

analyses with larger datasets lead to less sampling bias, one expects that larger datasets results in more variance explained. So, performing a traditional meta-analysis on the reduced dataset might have led to a lower percentage variance explained, thereby bringing the findings even more in line with each other.

Discussion

The central question addressed in this study is whether cognitive group differences represent true differences in general mental ability (*g*), or just "hollow" score differences. Scores on cognitive tests are the best general predictors of accomplishments in school and in the workplace, and it is predominantly the *g* component of IQ tests that is responsible for this criterion related validity. Combining the method of correlated vectors and psychometric meta-analysis the following theory emerges: When there is a correlation of +1.00 between the *g* vector and a second vector, variation in scores on the variable is caused by biological factors. When the correlation is –1.00 the variation in scores on the variable is caused by non-biological factors. When the correlation is close to zero the variation in scores on the variable is caused by roughly comparable biological and non-biological factors. In sum, a link was hypothesized between *g* loadings and a dimension of biological causation versus non-biological causation.

Almost perfectly in line with expectations the correlations with the vector of *g* loadings were highly positive for inbreeding depression, and negative for visual and hearing impairment for younger children. The MAs showed true correlations of .84 on inbreeding (K = 4; total N = 1,783), -.72 on visual impairment in predominantly young children (K = 6; total N = 238), and -.69 on hearing impairment in predominantly young children (K = 2; total N = 1,287), all three confirming our hypotheses. Taken together, the findings of inbreeding, visual impairment, and hearing impairment increase the likelihood of the hypothesized link between *g* loadings and a dimension of biological causation. Moreover, the lack of support of the deprivation theory further increases the plausibility of biological causation of group differences.

Contrary to our hypothesis, the exploratory MA on schizophrenia showed an uncorrected correlation of -.50 (K = 5; total N = 315). The exploratory MA on epilepsy showed an uncorrected correlation of .44, which is mildly supportive of our hypothesis (K = 7; total N = 445). A possible explanation could be that brain injuries following epilepsy or schizophrenia often pertain to specific areas of the brain. This would imply that brain damage following epilepsy or schizophrenia possibly doesn't affect *g*.

Methodological experiment on Giftedness

In the experimental meta-analysis on gifted children composite scores instead of scores on subtest level were being compared. The fact that the simplified procedure yields almost identical results as the more extensive procedure encourages further development of this new technique. Using composite scores would allow a larger number of studies to be included in psychometric meta-analyses.

However, some degree of caution is warranted when using this simplified technique. The fact that the FS scores are comprised of V and P scores poses some questions on the methodological feasibility of using composite scores. The effects of interdependency among the variables on the reliability of this new technique should be a topic of further investigation.

Practical implications

The present study makes a strong empirical contribution to the important discussion as to which interventions raise *g* and which do not. IQ tests are important instruments for selection and placement. Consequently, in today's society low IQ scores may lead to placement in special education whereas high IQ scores may lead to the placement in advanced training programs. Compensatory education aimed to lower the IQ gap between disadvantaged children and advantaged children and success of compensatory education was measured by IQ gains and improvement in scholastic achievement. However, increases due to schooling show very little or no transfer to general intelligence, suggesting that the massive sums spent on such programs have little chance of success.

Limitations of the studies

The present studies have a number of limitations. First, our meta-analyses are strongly based on the method of correlated vectors (MCV), and recently it has been shown to have limitations. Dolan and Lubke (2001) showed that when comparing groups substantial positive vector correlations can be obtained even when groups differ not only on *g*, but also on factors uncorrelated with *g*. Similarly, Ashton and Lee (2005) argue that spurious correlations between vectors may arise due to the fact that *g* loadings of a subtest are sensitive to the nature of other subtests in a battery. In the present study a strong negative correlation between schizophrenia and *g* loadings was found while +1.00 was expected. A possible explanation could be that the effect on performance subtests is stronger than the effect on verbal subtests, which in turn would lead to a negative correlation. This would implicate that highly positive correlations do not necessarily represent strong biological effects. Notwithstanding these limitations the MCV continues to be a widely used tool in scientific research. The present study contributes to the discussion on the merit of the MCV by analyzing a large number of empirical studies. Second, following Hunter

and Schmidt (1990) we excluded all extreme outliers from meta-analysis. Considering the fact that our K values range from four to seven, excluding extreme outliers means excluding a considerable percentage of the pertaining total sample size. Third, the MCV was applied using Pearson *r* which has the advantage of comparing the outcomes to earlier conducted meta-analyses also using Pearson *r*. However, we did not investigate whether the use of Spearman's rho would alter the robustness of the method. Future research should conduct meta-analyses using Spearman's rho and Pearson *r* separately in order to compare the results. Fourth, the majority of studies had no control group. Therefore our meta-analyses are based primarily on comparison between the focal group, and a nationally representative sample of the pertaining IQ test.

Conclusion

Based on a large number of empirical studies and employing the method of correlated vectors we developed a data-driven theory of a link between *g* loadings and a dimension of biological causation versus non-biological causation. This thesis added two full and two exploratory psychometric MAs yielding a total number of fourteen individual meta-analytical studies supporting the empirical basis of the theory and thereby increasing its plausibility.

References

References marked with an asterisk indicate studies included in the meta-analyses and in the exploratory studies.

*Adachi, N., Kanemoto, K., Muramatsu, R., Kato, M., Akanuma, N., Ito, M., Kawasaki, J., & Onuma, T. (2005). Intellectual prognosis of status epilecticus in adult epilepsy patients: Analysis with Wechsler Adult Intelligence Scale-Revised. *Epilepsia, 46,* 1502-1509.

*Afzal, M. (1988). Consequences of consanguinity on cognitive behavior. *Behavior Genetics, 18,* 583-594.

Alcock, K. J., & Bundy, D. A. P. (2001). The impact of infectious disease on cognitive development. In R. J. Sternberg & E. L. Grigorenko (Eds.), *Environmental effects on cognitive abilities.* Mahwah, NJ: Erlbaum.

*Anderson, R. J., & Sisco, F. H. (1977). Standardization of the WISC-R Performance scale for deaf children. (Office of Demographic Studies Publication Series T, No. 1). Washington, DC: Galloudet College.

Anderson-Roswall, L., Engman, E., Samuelsson, H., Sjöberg-Larsson, C., & Malmgren, K. (2004). Verbal memory decline and adverse effects on cognition in adult patients with pharmacoresistant partial epilepsy: A longitudinal controlled study of 36 patients. *Epilepsy and Behavior, 5,* 677-686.

Ashton, M. C., & Lee, K. (2005). Problems with the method of correlated vectors. *Intelligence, 33*, 431-444.

*Badaruddoza, & Afzal, M. (1993). Inbreeding depression and intelligence quotient among North Indian children. *Behavior Genetics, 23,* 343-347.

*Badaruddoza (2003). Effects of inbreeding on Wechsler intelligence test scores among North Indian children. *Asia Pacific Journal of Public Health, 16,* 99-103.

Bailey I. L., & Hall A. (1990). Visual Impairment: An overview. New York: American Foundation for the Blind.

*Baitinger, U., & Bernd, R. (1970). HAWIK-Untersuchungen bei Blinden und Sehhehinderten. *Der Blindenfreund,* 134-142.

Baker, G. A., Austin, N. A., & Downes, J. J. (2003). Validation of the Wechsler Memory Scale-III in a population of people with intractable temporal lobe epilepsy. *Epilepsy Research, 53,* 201-206

Barr, W. B. (1997). Receiver operating characteristic curve analysis of Wechsler Memory Scale-Revised scores in epilepsy surgery candidates. *Psychological Assessment, 9,* 171-176.

Bell, B. D. (2006). WMS-III Logical memory performance after a two week delay in temporal lobe epilepsy and control groups. *Journal of clinical and Experimental Neuropsychology, 28*, 1435-1443.

Berry, J.W. (1966). Temne and Eskimo perceptual skills. *International Journal of Psychology, 1*, 207-222.

*Bilder, R. M., Lipschutz-Broch, L., Reiter, G., Geisler, S. H., Mayerhoff, D. I., & Lieberman, J. A. (1992). Intellectual deficits in first-episode schizophrenia: evidence for progressive deterioration. *Schizophrenia Bulletin, 18, 3,* 437-448.

Bishop, V.E. (1996) *Teaching visually impaired children* (2nd Ed.). Springfield, IL: Charles C Thomas.

Bjørnæs, H., Stabell, K., Henriksen, O., & Løyning Y. (2001). The effects of refractory epilepsy on intellectual functioning in children and adults. A longitudinal study. *Seizure, 10,* 250-259.

Bourgeois, B. F. D., Prensky, A. L., Palkes, H. S., Talent, B. K., & Busch, S. G. (2004). Intelligence in epilepsy: A prospective study in children. Annals of Neurology, 14, 438-444.

*Braden, J. P. (1984). The factorial similarity of the WISC-R Performance scale in deaf and hearing samples. *Personality and Individual Differences, 5,* 403-409.

Braden, J. P. (1989). Fact or artifact? An empirical test of Spearman's hypothesis. *Intelligence, 13*, 149-155.

*Brown, S. W., & Yakimowski, M. E. (1987). Intelligence scores of gifted students on the WISC-R. *Gifted Child Quarterly, 31,* 130-134.

*Brown, S. W., Hwang, M. T., Baron, M., & Yakimowski, M. E. (1991). Factor analysis of responses to the WISC-R for gifted children. *Psychological Reports, 69,* 99-107.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.

Caruso, D. R., Taylor, J. J., & Detterman, D. K. (1982). Intelligence research and intelligent policy. In D.K. Detterman & R.J. Sternberg (Eds.), *How and how much can intelligence be increased* (pp. 45-65). Norwood, NJ: Ablex Publishing Corporation.

Christian, K., Bachnan, H. J, & Morrison, F. J. (2001). Schooling and cognitive development. In R. J. Sternberg & E. L. Grigorenko (Eds.), *Environmental effects on cognitive abilities*

(pp. 287-335). Mahwah, NJ: Erlbaum.
*Cohen, T., Block, N., Flum, Y., Kadar, M., & Goldschmidt, E. (1963). School attainments in an immigrant village. In E. Goldschmidt (Ed) The genetics of migrant and isolate populations. Baltimore: Williams & Wilkins.
Colom, R., Juan-Espinosa, M., & García, L. F. (2001). Secular increase in test scores is a "Jensen effect". *Personality and Individual Differences, 30*, 553-559.
Colom, R., Juan-Espinosa, M., Abad, F., & Garcia, L.F. (2000). Negligible sex differences in general intelligence. *Intelligence, 28,* 57-68.
Colom, R., Jung, R. E., & Haier, R. J. (2006). Distributed brain sites for the *g*-factor of intelligence. *NeuroImage, 31*, 1359-1365.
*Daugherty, K. M., & Moran, M. F. (1982). Neuropsychological, learning and developmental characteristics of the low vision child. *Visual Impairment and Blindness, 76,* 398-406.
Dekker, R. (1993). Visually impaired children and haptic intelligence test scores: Intelligence test for visually impaired children (ITVIC). *Developmental Medicine and Child Neurology, 35,* 478-489.
*Di Nuovo, S. F., & Buono, S. (2007). Psychiatric syndromes comorbid with mental retardation: differences in cognitive and adaptive skills. *Journal of Psychiatric Research, 41,* 795-800).
Dolan, C. V., & Lubke, G. (2001). Viewing Spearman's hypothesis from the perspective of multigroup PCA: A comment on Schonemann's criticism. *Intelligence, 29,* 231−245.
Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance. Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47,* 273−305.
Eysenck, H. J., & Barrett, P. (1985). Psychophysiology and the measurement of intelligence. In C. R. Reynolds & P. C. Wilson (Eds.), *Methodological and statistical advances in the study of individual differences* (pp. 1-49). New York: Plenum.
Fisher, I. (1930). The theory of interest.
Flynn, J. R. (1999a). Evidence against Rushton: The genetic loading of WISC-R subtests and the causes of between-group IQ differences. *Personality and Individual Differences, 26*, 373-379.
Flynn, J. R. (1999b). Reply to Rushton: A gang of *g*s overpowers factor analysis. *Personality and Individual Differences, 26*, 391-393.
Flynn, J. R. (2000). IQ gains, WISC subtests, and fluid *g*: *g* theory and the relevance of Spearman's hypothesis to race. In G. B. J. Goode (Ed.), *The nature of intelligence* (pp.

202-227). New York: Wiley.

Flynn, J. R. (2006). Efeito Flynn: Repensando a inteligência e seus efeitos [The Flynn Effect: Rethinking intelligence and what effects it]. In C. Flores-Mendoza & R. Colom (Eds.), *Introdução à Psicologia das Diferenças Individuais* (pp. 387-411). Brazil, Porto Alegre: ArtMed.

Gottfredson, L. S. (1997). Why *g* matters: The complexity of everyday life. *Intelligence, 24*, 79-132.

*Goldberg, T. E., Ragland, J. D., Torrey, E. F., Gold, J. M., Bigelow, L. B., & Weinberger, D. R. (1990). Neuropsychological assessment of monozygotic twins discordant for schizophrenia. *Archives of General Psychiatry, 47,* 1066-1072.

Glosser, G., Cole, L. C., French, J. A., Saykin, A. J., & Sperling, M. R. (1997). Predictors of intellectual performance in adults with intractable temporal lobe epilepsy. *Journal of the International Neuropsychological Society, 3*, 252-259.

Grantham-McGregor, S., Ani, C., & Fernald, L. (2001). The role of nutrition in intellectual development. In R. J. Sternberg & E. L. Grigorenko (Eds.), *Environmental effects on cognitive abilities.* Mahwah, NJ: Erlbaum.

Haier, R. J., Siegel, B., Tang, C., Abel, L., & Buchsbaum, M. S. (1992). Intelligence and changes in regional cerebral glucose metabolic rate following learning. *Intelligence, 16*, 415-426.

Hauser, W. A., & Hesdorffer, D. C. (1990). Epilepsy: Frequency, causes, and consequences. New York: Epilepsy Foundation of America.

*Henry, S. A., & Wittman, R. D. (1981). Diagnostic implications of Bannatyne's recategorized WISC-R scores for identifying learning disabled children. *Journal of Learning Disabilities, 14*, 517-520.

*Hiroshen, A., Hurley, O. L., & Kavale, K. (1979). Psychometric characteristics of the WISC-R Performance scale with deaf children. *Journal of Speech and Hearing Disorders, 44,* 73-79.

Hunter, J. E., & Schmidt, F. L. (1990). Methods of meta-analysis. London: Sage.

Hunter, J. E., & Schmidt, F. L. (2004). Methods of meta-analysis (2nd ed.). London: Sage.

Isham, W. P., & Kamin, L. J. (1993). Blackness, deafness, IQ, and *g. Intelligence, 17,* 37-46.

Jensen, A. R. (1974). Cumulative deficit: A testable hypothesis? *Developmental Psychology, 10,* 996-1019.

Jensen, A. R. (1978). Genetic and behavioral effects of nonrandom mating. In R. T. Osborne, C. E. Noble, and N. Weyl (Eds.), Human variation: The biopsychology

of age, race, and sex (pp. 51-105). New York: Academic Press.

Jensen, A. R. (1980). Bias in mental testing. New York: Free Press.

Jensen, A. R. (1983). Effects of inbreeding on mental ability factors. Personality and Individual Differences, 4, 71-87.

Jensen, A. R. (1994). Psychometric *g* related to differences in head size. *Personality and Individual Differences, 17*, 597-606.

Jensen, A. R. (1998). The g factor: The science of mental ability. London: Praeger.

Jensen, A. R., & Weng, L. -J. (1994). What is a good *g*? *Intelligence, 18*, 231-258.

Kamin, L. J. (1980). Inbreeding depression and IQ. *Psychological Bulletin, 87*, 469-478.

*Karnes, F. A., & Brown, K. E. (1980). Factor analysis of the WISC-R for the gifted. *Journal of Educational Psychology, 72,* 197-199.

Kim, H., Yi, S., Son, E. I., & Kim, J. (2003). Differential effects of left versus right mesial temporal lobe epilepsy on Wechsler Intelligence factors. *Neuropsychology, 4,* 556-565.

*Klauer, K. J. (1962). Sehschwäche und Intelligenz. *Zeitschrift für experimentelle und angewandte Psychologie.*

Klove, H., & Matthews, C. G. (1966). Psychometric and adaptive abilities in epilepsy with differential etiology. *Epilepsia, 7,* 330-338.

*Krüger, H. (1974). Intelligenzdiagnostik bei Sehbehinderten. *Der Blindenfreund,* 110-127.

Kulik, J. A., Bangert-Drowns, R. L., & Kulik, C. C. (1984) Effectiveness of coaching for aptitude tests. *Psychological Bulletin, 95,* 179-188.

Lee, K. H., Choi, Y. Y., Gray, J. R., Cho, S. H., Chae, J.-H., Lee, S., & Kim, K. (2006). Neural correlates of superior intelligence: Stronger recruitment of posterior parietal cortex. *NeuroImage, 29*, 578-586.

Lowenfeld, B. (1973). *The visually handicapped child in school*. London: Constable.

Lynn, R. (1982). IQ in Japan and the United States shows a growing disparity. *Nature, 297*, 222-223.

Lynn, R. (1997). Geographical variation in intelligence. In H. Nyborg (Ed.), *The scientific study of human nature*. Oxford: Pergamon.

Lynn, R. & Vanhanen, T. (2002). *IQ and the wealth of nations*. Westport, CT: Praeger.

*Macmann, G. M., Plasket, C. M., Barnett, D. W., & Siler, F. S. (1991). Factor structure of the WISC-R of superior intelligence. *Journal of School Psychology, 29,* 19-36.

Mommers. J. C., & Smits, B. W. G. M. (1975). Vergelijking van de prestaties van blinde en ziende kinderen op het verbale gedeelte van de WISC. *Pedagogische Studiën, 52,* 165-172.

Moore, P. M., & Baker, G. A. (1996). Validation of the Wechsler Memory Scale-Revised in a sample of people with intractable temporal lobe epilepsy. *Epilepsia, 37*, 1215-1220.

*Morice, R. (1990). Cognitive inflexibility and pre-frontal dysfunction in schizophrenia and Mania. *British Journal of Psychiatry, 157,* 50-54.

Must, O., Must, A., & Raudik, V. (2003). The secular rise in IQs: In Estonia the Flynn effect is not a Jensen effect. *Intelligence*, *167*, 1-11.

*Nelson, H. E., Pantelis, C., Carruthers, K., Speller, J., Baxendale, S., & Barnes, T. R. E. (1990). Cognitive functioning and symptomatology in chronic schizophrenia. *Psychological Medicine, 20,* 357-365.

Nelson, G., Westhues, A., & MacLeod, J. (2003). A meta-analysis of longitudinal research on preschool prevention programs for children. *Prevention & Treatment, 6*.

Nwuga, V. C. B. (1977). Effect of severe kwashiorkor on intellectual development among Nigerian children. *American Journal of Clinical Nutrition, 30*, 1423-1430.

Ogbu, J.U. (1994). Culture and intelligence. In R.J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 328-338). New York: Macmillan.

O'leary, S. D., Burns, T. G., & Borden, K. A. (2006). Performance of children with epilepsy ad normal age-matched controls on the WISC-III. *Child Neuropsychology, 12,* 173-180.

Prokosch, M. D., Yeo, R. A., & Miller, G. F. (2005). Intelligence tests with higher *g*-loadings show higher correlations with body symmetry: Evidence for a general fitness factor mediated by developmental stability. *Intelligence, 33*, 203-213.

Rath, W. (1967). Untersuchungen bei Blinden und hochgradig Sehbehinderten zum vergleich zwischen HAWIK-Verbalteil und IPT. *Der Blindenfreund,* 113-129.

Rae, C., Scott, R. B., Thompson, C. H., Kemp, G. J., Dumughn, I., Styles, P., Tracey, I., & Radda, G. K. (1996). Is pH a biochemical marker of IQ? *Proceedings of the Royal Society* (London), *263*, 1061-1064.

Ramey, C. T., Bryant, D. M., & Suarez, T. M. (1985). Preschool compensatory education and the modifiability of intelligence: A critical review. In D. Detterman (Ed.), *Current topics in human intelligence*. Norwood, NJ: Ablex Publishing Company.

*Reams, R., Chamrad, D., & Robinson, N. (1990). The race is not necessarily to the swift: Validity of WISC-R bonus points for speed. *Gifted Child Quarterly, 34,* 108-110.

Richardson, S. A., Birch, H. G., & Hertzig, M. (1973). School performance of children who were severely malnourished in infancy. *American Journal of Mental Deficiency, 77*, 623-632.

Sabatino, D. A., Spangler, R. S., & Vance, H. B. (1995). The relationship between the Wechsler Intelligence Scale for Children-Revised and the Wechsler Intelligence Scale for Children-III scales and subtests with gifted children. *Psychology in the Schools, 32,* 18-23.

*Sapp, G. L., Chissom, B., & Graham, E. (1985). Factor analysis of the WISC-R for gifted students: A replication and comparison. *Psychological Reports, 57,* 947-951.

Schafer, E. W. P. (1985). Neural adaptability: A biological determinant of *g* factor intelligence. *Behavioral and Brain Sciences, 8*, 240-241.

Schmidt, F. L., & Hunter, J. E. (1999). Theory testing and measurement error. *Intelligence, 27*, 183-198.

Schmidt, F. L., & Le, H. (2004). Software for the Hunter-Schmidt meta-analysis methods. University of Iowa, Department of Management & Organization, Iowa City, IQ, 42242.

Schneider, S. K., Nowack, W. J., Fitzgerald, J. A., Janati, A., & Souheaver, G. T. (1993). WAIS performance in epileptics with unilateral interictal EEG abnormalities. *Journal of Epilepsy,* 6, 10-14.

Schoenemann, P. T. (1997). The evolution of the human brain and its relationship to behavior. Unpublished doctoral dissertation. Berkeley: University of California.

*Schull, W. J., & Neel, J. V. (1965). The Effects of Inbreeding on Japanese Children. New York: Harper & Row.

*Sevier, R. C., & Bain, S. K. (1994). Comparison of WISC-R and WISC-III for gifted students. *Roeper Review, 17,* 39-42.

Singhi, P. D., Bansal, U., Singhi, S., & Pershad, D. (1992). Determinants of IQ profile in children with idiopathic generalized epilepsy. *Epilepsia, 33,* 1106-1114.

Shuey, A.M. (1966). *The testing of Negro intelligence* (2nd Ed.). New York: Social Science Press.

*Spangler, R. S. & Sabatino, D. A. (1995). Temporal stability of gifted children's intelligence. *Roeper Review, 17,* 207-210.

Stella, F., & Maciel, J. A. (2004). Intelligence functions disorders in patients with complex partial epilepsy. *Arquivos de Neuropsiquiatria, 62,* 983-987.

Stephens, B., & Grube, C. (1982). Development of Piagetian reasoning in congenitally blind children. *Visual Impairment and Blindness, 76,* 133-143.

*Studeny, P. (2008). Cognitive impairment of blind and visually impaired children in Austria and South Africa. Kognitive Fähigkeiten blinder und sehbehinderten Kinder im Ländervergleich Österreich und Südafrika. *Master's thesis, Karl-Franzens university, Graz, Austria.*

Swanson, H. E., & Lussier, C. M. (2001). A selective synthesis of the experimental literature on dynamic assessment. Review of Educational Research, 71, 321−363.

te Nijenhuis, J., & Jongeneel-Grimen, B. (2007). *Can people get smarter?: Three psychometric meta-analyses and three exploratory meta-analyses on g loadings and IQ scores*. Unpublished manuscript, University of Amsterdam, Amsterdam, NL.

te Nijenhuis, J., Tomic, W., & Franssen, D.B. (2008). Is there a linear relationship between inbreeding and mental ability?: A meta-analysis. Unpublished manuscript, University of Amsterdam, Amsterdam, NL.

te Nijenhuis, J., & van der Flier, H. (2007). The secular rise in IQs in the Netherlands: Is the Flynn effect on g? *Personality and Individual Differences, 43*, 1259-1265.

te Nijenhuis, J., & van der Flier, H. *Is the Flynn effect on g?: A meta-analysis*. Manuscript submitted for publication.

te Nijenhuis, J., van Rijk, R., & Kämper, S. (2007). Blindness, deprivation and IQ: A meta-analysis. Unpublished manuscript, University of Amsterdam, Amsterdam, NL.

te Nijenhuis, J., van Vianen, A. E. M., & van der Flier, H. (2007). Score gains on *g*-loaded tests: No g. *Intelligence, 35,* 283-300.

Tillman, H. M., & Osborne, R. T. (1969). The performance of blind and sighted children on the Wechsler intelligence scale for children; Interaction effects. *Education of Visually Handicapped, 1,* 1-4.

van Haasen, P. P., de Bruyn, E. E. J., Pijl, Y. J., Poortinga, Y. H., Lutje Spelberg, H. C., Vander Steene, G., Coetsier, P., Spoelders, R., & Stinissen, J. (1986). *WISC-R: Wechsler Intelligence Scale for Children-Revised; Nederlandstalige uitgave* [WISC-R: Wechsler Intelligence Scale for Children-Revised; Dutch edition]. Lisse, the Netherlands: Swets.

Vernon, P. A. (1992). Intelligence, reaction times, and peripheral nerve conduction velocity. *Intelligence, 16*, 273-288.

Vernon, P. A. (1993). Intelligence and neural efficiency. In D. K. Detterman (Ed.), *Current topics in human intelligence,* Vol. 3 (pp. 171-187). Norwood, NJ: Ablex.

*Wechsler, D. (1958). *The measurement and appraisal of adult intelligence.* Baltimore: Williams & Wilkins.

Wechsler, D. (1974). Manual for the Wechsler Intelligence Scale for Children - Revised. New York: The Psychological Corporation.

*Wechsler, D. (2003). *Wechsler Intelligence Scale for Children-Fourth Edition.* San Antonio, TX: The Psychological Corporation.

*Wheaton, P. J., & Vandergriff, A. F. (1978). Comparison of WISC and WISC-R scores of highly gifted students in public school. *Psychological Reports*, *43*, 627-630.

Wicherts, J. W., Dolan, C. V., Oosterveld, P., van Baal, G. C. V., Boomsma, D. I., & Span, M. M. (2004). Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect. *Intelligence, 32*, 509-537.

Wickett, J. C., Vernon, P. A., & Lee, D. H. (1994). In *vivo* brain size, head perimeter, and intelligence in a sample of healthy adult females. *Personality and Individual Differences, 16*, 831-838.

*Wilkinson, S. C. (1993). WISC-R profiles of children with superior intellectual ability. *Gifted Child Quarterly, 37,* 84-91.

Wright, S.C., Taylor, D.M., & Ruggiero, V.R. (1996). Examining the potential for academic achievement among Inuit children: Comparisons on the Raven coloured progressive matrices. *Journal of Cross-Cultural Psychology, 27*, 733-753.

Zeidner, M. (1987). Test of the cultural bias hypothesis: Some Israeli findings. *Journal of Applied Psychology, 72*, 38-48.